

# Psychological Review

THEODORE M. NEWCOMB, Editor  
*University of Michigan*

---

## CONTENTS

Uncertainty and Conflict: A Point of Contact Between  
Information-Theory and Behavior-Theory Concepts.....D. E. BERLYNE 329

Neural Mechanisms in Perception.....JEROME S. BRUNER 340

Motivational Determinants of  
Risk-Taking Behavior.....JOHN W. ATKINSON 359

Durable Secondary Reinforcement: Method and  
Theory.....DONALD W. ZIMMERMAN 373

An Opponent-Process Theory of  
Color Vision.....LEO M. HURVICH AND DOROTHEA JAMESON 384

---

This is the last issue of Volume 64.  
Title page and index for the volume  
appear herein.

---

PUBLISHED BIMONTHLY BY THE  
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

## Consulting Editors

SOLOMON E. ASCH  
ROBERT R. BLAKE  
STUART W. COOK  
CLYDE H. COOMBS  
WILLIAM K. ESTES  
LEON FESTINGER  
W. R. GARNER  
JAMES J. GIBSON  
D. O. HEBB  
HARRY HELSON  
E. R. HILGARD  
CARL I. HOVLAND  
E. LOWELL KELLY  
DAVID KRECH  
ROBERT W. LEEPER  
KENNETH MACCORQUODALE

ROBERT B. MACLEOD  
DAVID C. MCCLELLAND  
GEORGE A. MILLER  
FREDERICK MOSTELLER  
GARDNER MURPHY  
OSCAR OESER  
CARL PFAFFMANN  
CARROLL C. PRATT  
JOHN P. SEWARD  
DAVID SHAKOW  
RICHARD L. SOLOMON  
ELIOT STELLAR  
S. S. STEVENS  
ERIC TRIST  
EDWARD L. WALKER  
ROBERT W. WHITE

ARTHUR C. HOFFMAN, Managing Editor

HELEN ORR, Assistant Managing Editor

Editorial Staff: SADIE J. DOYLE, FRANCES H. CLARK, BARBARA CUMMINGS,  
SARAH WOMACK

The *Psychological Review* is devoted to theoretical articles of significance to any area of psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding twelve printed pages (about 7,500 words) are not accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail. Manuscripts should be addressed to the Editor, Dr. Theodore M. Newcomb, Doctoral Program in Social Psychology, University of Michigan, Ann Arbor, Michigan.

PUBLISHED BIMONTHLY BY THE  
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
PRINCE AND LEMON STREETS, LANCASTER, PA.  
1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

\$2.00 volume

\$1.50 issue

Entered as second-class matter July 13, 1897, at the post office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 3640,

P. L. & R. of 1948, authorized Jan. 8, 1948

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1957 by the American Psychological Association, Inc.

# THE PSYCHOLOGICAL REVIEW

## UNCERTAINTY AND CONFLICT: A POINT OF CONTACT BETWEEN INFORMATION-THEORY AND BEHAVIOR-THEORY CONCEPTS<sup>1</sup>

D. E. BERLYNE

*Center for Advanced Study in the Behavioral Sciences<sup>2</sup>*

*Information theory*, originally designed to handle certain problems in communications engineering (41), needs to be distinguished from *psychological information theory*, which is one of its offshoots. The former consists of a mathematical language, incorporating a number of distinctive measuring techniques. Psychological information theory is, in contrast, a type of theory in the scientific sense: it applies information-theory measures to phenomena within the purview of psychology and uses information-theory language to formulate laws or hypotheses with testable implications about behavior.

Recent literature contains several sketches of such theory (e.g., 1, 26, 35), mostly concerned with how human beings code information or with how much information can pass through them in particular situations. There are many unmistakable affinities between this kind of psychological theory and S-R behavior theory (learning theory): they have overlapping interests in

such matters as discrimination, remembering and reaction time, they share a predilection for operationally definable and quantitative concepts, and they start out, respectively, from the closely related "black-box" and "neobehaviorist" points of view. It is therefore rather disappointing that so little integration between theories of the two types has yet taken place. We can regard two theories as "integrated" if one can be deduced from the other or if both can be deduced from a third theory. Before any integration can be attempted, the scope of information-theory language within the domain of behavior theory must be examined, which means considering to what extent recourse to it is *possible* and to what extent, if possible, it is *useful*.

The use of information-theory measures is *possible* whenever we have a *partition*, i.e., a set of phenomena that can be divided into non-overlapping subsets, and a *probability distribution*, i.e., a way of associating with each subset a number from 0 to 1, such that the numbers associated with all the subsets in the partition add up to 1. Whenever these two requirements are fulfilled, such measures as "amount of information," "uncertainty," and "relative uncertainty" can be applied. As

<sup>1</sup> This article owes a great deal to discussions with Dr. I. R. Savage and several other colleagues at the Center for Advanced Study in the Behavioral Sciences. It was written while the author was on leave of absence from the University of Aberdeen, Scotland.

<sup>2</sup> Now at the University of California, Berkeley.

soon as we have *two* sets of phenomena satisfying both requirements, the two can jointly be regarded as a "transducer," and the relations between them described in terms of "transmitted information," "noise," and "equivocation."

The phenomena that concern behavior theory consist, in fact, of two sets that can be partitioned into subsets with associated probabilities, namely *stimuli* and *responses*. The language of information theory is therefore, in principle, applicable to everything within the competence of behavior theory. Two limitations to its appropriateness have often been pointed out (14, 47). One is that the behavior theorist is especially interested in learning, i.e., in situations where probabilities of responses are changing. The other is that information-theory measures take no account of any ordering of the subsets in the partition or, more particularly, of the fact that stimuli and responses are not confined to nominal scaling (44). But these limitations are not insuperable. Information-theory measures can be derived from response probabilities at different stages of a learning process and compared, or else they can be applied when learning is near its asymptote. Stimuli and responses can be successively described in information-theory terms and in terms of physical or psychophysical dimensions, and the different measures can be related. For example, a response-class can have attached to it both a mean reaction time and a rate of transmitted information, and connections between the two can be explored.

If the possibility of describing the domain of behavior theory in information-theory language is accepted, the question of its *usefulness* still remains. One of the principal functions of any language is to make secondary or mediated generalization and discrimination possible. A language incorporates classifications, of which measures are spe-

cial cases. Classifications are procedures for attaching certain descriptive terms (values in the case of measures) as verbal responses to certain items in the universe of discourse but not to others. Items bearing a common verbal label come to evoke similar behavior in the users of the language. A classification is useful only as long as the items allotted the same label share some important quality, such that a common response to them will be rewarded (or reinforced) despite other qualities that might distinguish them. Information-theory measures are useful for the description of behavior, therefore, if these measures are closely related to other variables that have proved to be important for psychology.

A large body of data demonstrating that such is the case has been amassed within the last ten years. Reaction time, retention of verbal material, and accuracy of psychophysical judgment, to cite examples, appear to be functions of "uncertainty" and "amount of transmitted information." The situations in which such associations have been found have, however, been situations in which *subjects have some knowledge of the range of alternative stimuli that might occur and of their probabilities*. This knowledge is provided by *E's* instructions, or by the presentation of a sufficient sample of material for estimates to be made, or, as in experiments using natural languages, by previous training. It has, indeed, been contended by Cronbach (14) that information-theory measures in psychology should be confined to cases where "the receiver knows the probabilities and joint probabilities of the source." As Cherry reminds us, information theory is part of the "meta-language of an external observer; it is not a description of the process of communication as it appears to one of the participants" (13, p. 170). An observer can compute information-theory meas-



ures from data not accessible to the individuals he is observing. But there is not likely to be much connection between these measures and variables of psychological importance, unless there is some isomorphism between the situation as viewed by the observer and the situation as it impinges on the observed organism.

The situations in which the use of information-theory terminology has had some success can be analyzed further as follows:

1. There is an antecedent stimulus-pattern,  $S_a$ . It may consist of the background conditions of the experiment, of an  $E$ 's warning signal or, in sequential studies, of any item in a sequence.

2. Whenever  $S_a$  occurs, it is followed by one and only one of a set of consequent stimuli  $\{S_1 \dots S_n\}$ .

3. Whenever one of the consequent stimuli occurs, a particular response corresponding to it is performed.

4. The responses corresponding to the consequent stimuli are such that no more than one of them can be performed at once, whether because of the  $E$ 's instructions or because of some physiological incompatibility between them.

In such situations, one can predict that all the  $n$  responses corresponding to the  $n$  consequent stimuli will become conditioned to  $S_a$ . No more than fractional components of these responses can be expected to occur immediately after the onset of  $S_a$ , both because simultaneous performance of the complete responses is precluded by the conditions of the experiment and because performance of any of them before the consequent stimulus appears will not be reinforced, so that the conditions for inhibition of delay will be fulfilled (38).  $S_a$  will thus come to evoke *competing response tendencies*. For Hull's theory

(27, 28), these response tendencies will be "reaction potentials." Cognitive behavior theories (e.g., 45) would describe them as "expectations" of the consequent stimuli, and the "expectation" resembles the "reaction potential" insofar as both imply the occurrence of a particular response, if certain additional conditions are met.

Furthermore, the relative *strengths* of the competing response tendencies will reflect the probabilities of the corresponding stimuli. Whether one regards the number of reinforced trials (27, 28), the variety of stimulus situations that have been contiguous with the response (23), or the number of times an expectation has been confirmed (45) as the decisive factor, responses associated with more frequent consequent stimuli will become more strongly associated with  $S_a$ . There is, in fact, experimental evidence (17, 21) that the strength of a predictive verbal response (which is especially relevant here), as judged by the asymptote of response probability, increases with the probability of the corresponding stimulus.

To sum up, the situations in which information-theory language has been of value are ones in which *conflict* is an important factor, and the theory of conflict seems to be one area where linkages between information theory and behavior theory may hopefully be sought.

#### DEGREE OF CONFLICT (C)

If the study of conflict is to progress beyond noting the effects of its presence or absence, some way of distinguishing *degrees of conflict* will have to be adopted. The degree of conflict is, of course, not necessarily the same as the *severity of the effects of conflict*, of which it is likely to be merely one determinant. Other determinants would be the nature of the conflicting response tendencies (e.g., whether they are approach or avoidance tendencies [36])

and the conflict tolerance of an individual organism. Brown and Farber (11) suggest two conditions for the degree-of-conflict function (or, as they call it, "frustration"), viz., that it increase with the *absolute strengths* of the competing tendencies, and that it increase as their strengths approach *equality*. They, like most writers who have considered psychological conflict, confine their attention to conflicts between two response tendencies. If conflicts involving three or more alternatives are to be included in the treatment, as would seem desirable, the *number* of competing tendencies can be proposed as a third variable with which degree of conflict increases (4).

We can thus state the conditions for a degree-of-conflict function a little more precisely, as follows. Let us assume that response tendencies corresponding to a set of responses  $\{R_1 \dots R_n\}$  occur in an organism, that the responses in the set are such that no two of them can be performed at once, and that some non-negative quantity  $E$  (e.g., Hull's "reaction potential") can be associated with each response tendency (as a measure of its strength).

It should be noted that, although the *responses* cannot occur simultaneously, we are assuming that their corresponding *response tendencies* can. Second, there is no reason why several independent sets of competing response tendencies should not be aroused in the same organism at once. Third, we are considering cases where there is complete incompatibility, whether innate or learned, between alternative responses. It is, however, conceivable that two responses may be *partially* antagonistic, i.e., the evocation of one may reduce the amplitude or probability of the other without excluding its performance altogether. This may suggest *degree of incompatibility* between responses as an additional determinant of degree of conflict (4), which would complicate any mathematical treatment. Possible ways

of reducing degree of incompatibility to other variables, when learned incompatibility is involved, are considered elsewhere (7).

The degree-of-conflict function  $C(E_1 \dots E_n)$  should then have the following properties:

1.  $C$  is continuous and symmetric in the  $E_i$ ;
2.  $C \geq 0$ ;
3. if  $n = 1$ ,  $C = 0$ ;
4. with  $\sum_{i=1}^n E_i$  held constant,  $C$  reaches an absolute maximum when  $E_1 = E_2 = \dots = E_n$ ;
5. if  $E_1 = E_2 = \dots = E_n$ , and a response  $R_{n+1}$  with strength  $E_{n+1} > 0$  is added to the set  $\{R_1 \dots R_n\}$ ,  $C$  increases;
6. if every  $E_i$  is multiplied by  $k > 1$ ,  $C$  increases.

Now, let us suppose that we have a way of translating the  $E$  value for each response into a  $p$  value, or measure of probability, such that

1.  $0 \leq p \leq 1$ ;
2.  $\sum_{i=1}^n p_i = 1$ ;
3. if  $E_1 = E_2 = \dots = E_n$ , then  $p_1 = p_2 = \dots = p_n = 1/n$ ;
4. if one  $E_i$  increases with the others held constant, then the corresponding  $p$  increases and the other  $p$ 's decrease.

Some theorists (e.g., 12, 16) content themselves with probability as a sole measure of response strength. Others (e.g., 27, 42, 43) recognize additional ones, such as latency, frequency, resistance to extinction, amplitude and vigor. Of these, mean latency and mean frequency are merely the reciprocal of the probability that a response of the class in question will occur during one unit of time. Resistance to extinction may be regarded as the rate at which response

probability decreases when reinforcement is withdrawn. But other measures of response strength, expressing the energy with which the response is performed, are not the same as probabilities. Hull (28, pp. 25 ff.) and Spence (43, App. A) present methods for transforming  $E$ s into probabilities, when  $E$  is the Hullian reaction potential. Doing this means, however, losing information, since many sets of  $E$  values can be represented by the same set of  $p$  values. Whenever we have two or more independently defined response classes, as distinct from one response class and its complement, probability is a measure of *relative* and not absolute response strength. The distinction may be important. For example, Mr. A. may be torn between his duty to the community and his duty to his family, while Mr. B may have difficulty in deciding whether or not to spend a small sum on a newspaper. Both of them have two response tendencies with probabilities of .5, but in other respects the effects of the two conflicts may be radically different.

Be that as it may, the use of probabilities to express response strengths provides us with the partition and the probability distribution that are necessary conditions for recourse to information-theory measures. And if we examine the information theorist's formula for "uncertainty" or "entropy" ( $-\sum_i p_i \log_2 p_i$ ), we find that it satisfies the first five of our requirements for a degree-of-conflict function, but not the sixth. It increases with the number of alternative responses and is at a maximum when their strengths are equal. But it does not vary with their absolute strengths. In order to make "uncertainty" fulfill all our conditions, we can multiply it by some such quantity as the mean  $E$ . French's hypothesis (20) that the frustrating effects of a binary conflict are a function of the weaker of the two opposing forces suggests that "uncertainty" should be multiplied by the minimum rather than the mean  $E$ . But this would produce rather anomalous

results in higher-order conflicts when there are one very weak and several very strong response tendencies in competition. Our expression for degree of conflict then becomes  $-\bar{E} \sum_i p_i \log p_i \cdots 1$ . Put somewhat differently, "uncertainty" can be regarded as an indication of the "complexity" of a conflict, or of the difficulty that an observer would have in predicting which of the conflicting responses will be the first to occur. It does not reflect the "scale" of the conflict, which depends on the energy invested in the competing response tendencies. There may be a temptation to relate these two components to the *utility* and *probability-of-outcome* factors that must be taken into account in decision theory, or to the *motivational* and *structural* factors that often have been distinguished in psychological literature. But any such correspondence would be misleading. Both the "uncertainty" and the  $\bar{E}$  are determined by absolute response strengths, which depend on both motivational (utility) and structural (probability-of-outcome) variables; e.g., Hull's "reaction potential" (28) depends on "drive" and "amount of reinforcement" on the one hand and on "number of reinforcements" (habit-strength) on the other. It is interesting to observe that Shannon (41, p. 19) gives  $-K \sum_i p_i \log p_i$  as the only function satisfying his assumptions, and goes on to describe  $K$  as amounting to a "choice of a unit of measure" or, in other words, to some scaling factor comparable to our  $\bar{E}$ .

Expression 1 is, however, by no means the only one that will accord with our requirements. Another function, for example, that will do so without necessitating a transformation of  $E$ , is  $\sum_i (\log (\sum_i E_i) - \log E_i) \cdots 2$ . If  $E$  represents Hull's reaction potential, this function will not, in general, have the same values as Expression 1, because probabilities are not proportional to reaction potentials. It will, however, be an increasing monotonic transform of Expression 1.

We are not even confined to logarithmic functions, since we lack the additivity requirement that makes them mandatory for Shannon's purposes. A non-logarithmic function that will pass muster is

$$\frac{(\sum E_i)^2(n-1)}{1 + \sum_i (E_i - \bar{E})^2} \dots 3.$$

Our requirements are, in fact, very weak ones, which a large number of functions will fit. Further research will, no doubt, add stipulations, allowing the range of possible functions to be narrowed down. For instance, one additional requirement that may be held reasonable, in view both of everyday observation of persons confronted with choices and of the logarithmic relation that obtains between number of alternative stimuli and choice reaction time, is that  $C$  should be a negatively accelerated increasing function of  $n$ . If this were adopted, then Expression 1 would be among those still meriting consideration, but Expressions 2 and 3 would be ruled out.

#### CORRELATES OF DEGREE OF CONFLICT

A degree-of-conflict measure, like an information-theory measure, can be justified as a classificatory device only if situations that have a common value assigned to them by the measure result in similar behavior, much as they may differ in other respects. The following are some psychological variables that appear likely, in the light of present knowledge, to depend on degree of conflict. They may actually turn out to be closely interrelated, but they are here separated for convenience.

1. *Emotional disturbance.* Various writers, from Dewey (15) on, have mentioned conflict as a cause of "emotion." Both the special reaction patterns (24) and the disruption of habitual behavior (31) that are characteristic of "emotional disturbance" have been ascribed to the occurrence of divergent neural processes. The power of conflict to precipitate neurotic behavior was pointed out independently by Pavlov and by

Freud, employing very different research techniques. So far, merely the dependence of these phenomena on relatively severe conflict has been noted, but future progress may well demand a quantitative treatment, in which intensities of disturbance are differentiated and related to degrees of conflict.

2. *Reaction time.* A lengthening of reaction time (or decision time or choice time) has often been reported as a consequence of conflict (see Berlyne [7]). A link with information theory presents itself in the finding that reaction time increases linearly, at least in some conditions, with "uncertainty" (26, 29); mean reaction time has been found to increase when alternative stimuli approach equiprobability and when they become more numerous. If, as we concluded, the number of competing response tendencies corresponds to the number of alternative stimuli, and if the relative strengths of those tendencies reflect the probabilities of the corresponding stimuli, we can infer that two of the suggested determinants of  $C$  affect reaction time.

Both traditional experimental psychology and psychological information theory have hitherto concentrated on "forced-choice" situations, in which only one response is appropriate to each alternative stimulus, and selection of a response depends on identification of the stimulus. A recent investigation by the writer (7) compared forced choices with *free choices*. For the latter, two or more stimuli were presented together, and the response corresponding to any one of them was to be performed. Both kinds of choice can be assumed to entail conflict: the free choice means a conflict between response tendencies of about equal strength evoked by the stimuli that are simultaneously present, while the forced choice means an unequal and therefore relatively mild conflict between a strong tendency to respond correctly to the one stimulus that occurs and weak tendencies to make re-

sponses appropriate to other stimuli, resulting from generalization. The usual information-theory analysis of the forced choice, in which the *S* is viewed as a transducer with a limited channel capacity, is not helpful for the treatment of the free choice.

Free-choice reaction times invariably exceeded forced-choice reaction times, and both were longer when the number of alternative stimuli and responses was increased from two to four, as the hypothesis that reaction time increases with degree of conflict would lead one to expect. Furthermore, when the absolute strengths of the response tendencies—the determinant of degree of conflict that is disregarded by uncertainty—were manipulated by changing the intensity or extensity of the stimuli, changes in free-choice reaction time resulted.

3. *Drive*. Various considerations and observed phenomena have led a number of writers (e.g., 11, 30, 46) to conclude that conflict may be a drive condition. The drive resulting from conflict as such must, of course, be distinguished from other drives that may be at a high level because conflict blocks the behavior that would normally reduce them.

A certain amount of evidence for a conflict drive was obtained by Lowell (30), who found approach-approach conflict to produce a greater speed of running in rats than a single approach tendency. A supplementary observation fitting our conception of *C* was that the conflict drive was not so much in evidence when the stimuli were unequal in intensity or when learning was incomplete (and the competing response tendencies presumably relatively weak).

Wyckoff's experiment (48, 49) provides other data that might be predicted from our assumptions. His pigeon *Ss* were rewarded with food when they pecked at a key of a certain color and not rewarded when the key was of another color. They were then tested with the key white, but the color indi-

cating whether pecking would be reinforced or not appeared if the animal stepped on a pedal. The pedal response was rapidly learned, even though it did not affect the probability of receiving food. It merely diminished the pigeon's "uncertainty" by one bit. The white key is reminiscent of the stimulus that made the dog neurotic in the famous Shenger-Krestovnikova experiment (38, pp. 290 ff.). This stimulus, intermediate in shape between the reinforced circle and the nonreinforced ellipse, was thought by Pavlov (38, p. 318), to produce a "conflict between excitation and inhibition." If the white key produced a conflict in Wyckoff's pigeons between tendencies to peck and to refrain from pecking, or between tendencies to expect and not to expect food, the coloring of the key that was a consequence of stepping on the pedal must have reduced the conflict by strengthening one response tendency and weakening the other. If a conflict drive is proportional to *C*, reduction of the drive can be expected to reinforce the pedal-stepping response. When the discrimination was reversed, Wyckoff found that the frequency of the pedal response would temporarily decrease. This also fits our interpretation, as each color would then go through a stage of evoking both tendencies, and seeing the colored key would thus increase rather than reduce conflict. Wyckoff himself offers an alternative explanation in terms of secondary reinforcement, but this leads into difficulties, as Prokasy points out in his report of a somewhat similar experiment (39).

Yet another relevant experiment is one by Fonberg (19). She trained dogs to perform a certain response (*R*<sub>1</sub>) as a way of terminating stimuli that had been associated with puffs of air or electric shocks. The animals then received training in quite a different response (*R*<sub>2</sub>), which was followed by food reinforcement in the presence of a loud tone but not in the presence of a



faint tone. When they were later subjected to a Shenger-Krestovnikova type of conflict by exposure to tones intermediate in intensity between the positive and negative alimentary conditioned stimuli, they reverted to their defensive response ( $R_1$ ). This finding indicates that the physiological state produced by a conflict, even when noxious stimuli have played no part in it, may be sufficiently similar to the physiological state (fear or anxiety) resulting from a noxious stimulus for generalization between the two to occur.

4. *Curiosity*. There is currently a good deal of interest in certain sorts of behavior whose main function seems to be the provision of information, and information theory might reasonably be expected to throw some light on them. The behavior under discussion includes the "exploratory" activities that bring about opportunities to perceive objects more readily; the verbal activities, including asking questions, that elicit informative verbal behavior from other individuals; and the symbolic activities that allow thought processes to feed on information other than that supplied by the immediate environment.

"Novelty" has often been mentioned as a distinguishing mark of situations that provoke such activities (2, 6). But something can either be relatively novel, in the sense that it has never been encountered before in its present context, or absolutely novel, in the sense that it has never been encountered at all. In both cases, we have situations in which "amount of information" is high, since this measure is inversely related to the probability of an event, and the probability of particular novel occurrence must be low in the light of an individual's past experience. We can also speak of conflict in connection with the same occurrences. A relatively novel stimulus pattern is one in which perception conflicts with the expectations aroused by the context. Moreover, at least as far as human beings are con-

cerned, any absolutely novel object is bound to consist of an unfamiliar combination of familiar elements or to possess characteristics intermediate between those of several well-known objects. Such an object can be expected to induce conflict, since it will inevitably evoke, by generalization, responses appropriate to a number of discrepant familiar objects.

Other words that seem apposite to situations that call for investigatory behavior are "doubt," "perplexity," and "ambiguity." These words likewise imply some degree of behavioral conflict; they indicate that different aspects of a situation evoke discordant reactions or else that a particular reaction is called forth by one aspect and inhibited by another. They are opposite in meaning to words like "clear" and "distinct," which generally imply that certain response tendencies have come, through discriminatory learning, to predominate over their competitors. "Doubtful," "perplexing," or "ambiguous" stimulus situations are usually also cases of high "uncertainty" in the information-theory sense, both because the subject cannot predict very successfully what the future behavior or the hidden properties of the entities will be, and because observers will not be able to predict very successfully how he will react to them. Nevertheless, curiosity is by no means always commensurate with "uncertainty"; there are many events whose outcomes are uncertain and yet which leave us completely indifferent. For knowledge of the outcome to be rewarding, the event must be of some "interest" to us, which usually means that strong habits or drives must be aroused. In other words, curiosity seems to be a matter of conflict rather than of "uncertainty" alone; "uncertainty" may be high, but there will not be much conflict if the absolute response strengths are low. That human beings, like Wyckoff's pigeons, find relief from doubt about vital matters rewarding, even when the

truth is unpleasant, is attested by common experience. Of the convicts studied by Farber (18), those who did not know how much time they would have to serve suffered more than those who were certain that they would never be paroled.

The writer suggested a few years ago (4) that at least some forms of human curiosity spring from the drive-producing properties of conflict. The conflicts that seem especially pertinent are those between implicit, most often symbolic, responses, such as "beliefs," "ways of thinking," and "ways of perceiving," whose incompatibility is largely an effect of learning. There are experimental data supporting the conclusion that curiosity, measured in various ways, is an increasing function of  $C$  (5, 8, 9).

5. *Stimulus complexity.* Among the various properties by which stimulus patterns can be classified, there is a group that can only be described collectively by some such term as "complexity." They are hard to define rigorously, and a number of quite distinct dimensions will, in all likelihood, be unraveled by attempts to do so. But the influence of this aspect of perceived material is revealed in several contexts: the special properties attributed to less complex (more "prägnant") figures by the Gestalt school, the bearing of degree of complexity on aesthetic preferences, and, more recently, the influence of stimulus complexity on exploratory behavior on animals (see 10).

Attneave (1) has related the "complexity" dimension in visual figures to information theory through the concept of "redundancy," the inverse of "relative uncertainty." His treatment suggests a possible link between these variables and conflict. More "complex" stimulus patterns might well be those arousing more conflict, e.g., between perception of one part and expectations or redintegrative perceptual responses (3, 24, 37) aroused by other parts, between verbal or other classificatory re-

sponses, or between ocular and other orienting movements. If this hypothesis is well founded, we should expect more "complex" (or less "redundant") figures, like figures arousing conflict in other ways, to elicit more investigatory behavior. Experimental data confirming this prediction are available (8, 9).

6. *Reward.* While the punishing or drive-producing role of conflict is more evident and has received more attention, the possibility that conflict and uncertainty may at times be rewarding is suggested by gambling and aesthetic behavior. Similarly, journalistic practice seems to indicate a positive relation between the reward value of a piece of news and the "amount of information" it contains, which depends on its improbability or surprisingness (40). Surprise, like novelty, seems to mean some sort of clash between the reactions occasioned by an unexpected situation and those evoked anticipatorily through previously established habits (3). Surprising statements are, at least in certain circumstances, recalled more readily than others (5), and maze-learning experiments (see 10) show that exposure to a more complex environment (which, as we have seen, may mean a more conflictful environment) can be more reinforcing than exposure to a simpler one.

If conflict is usually an aversive condition but occasionally functions as a reward, it resembles fear, which likewise seems to be actively sought at times, e.g., at fairgrounds and in dangerous sports. The analogy with fear suggests two hypotheses to account for the paradox. One is that drive arousal may be rewarding at a moderate level. Hebb refers to "the positive attraction of risk taking, or mild fear, and of problem solving, or mild frustration," and speculates that "when arousal or drive is at a low level, . . . a response that produces increased stimulation and greater arousal will tend to be repeated" (25, p. 250). McClelland *et al.* (33) propound a rather similar hypothesis, whose bear-

ing on conflict is a little more conspicuous: "positive affect is the result of smaller discrepancies of a sensory or perceptual event from the adaptation of the organism; negative affect is the result of larger discrepancies." There have been a number of recent studies (e.g., Marx *et al.* [32]) showing that an increase in illumination up to a certain intensity will reinforce a bar-pressing response in a rat, while light of much greater intensity is known to be aversive.

The second hypothesis is that such states as fear or conflict are sought only when their arousal in similar circumstances has reliably and speedily been followed by drive reduction in the past.

Two recent works by empirically minded aestheticians provide some corroboration for these hypotheses. Graves (22) contends that the appeal of a visual design depends on variety, but that one part or quality must be made to dominate the others if the effect is to be satisfying. This would keep within bounds any conflict aroused. In accord with the second hypothesis, Meyer (34) shows that music owes much of its savor to continual departures from what preceding or accompanying patterns lead the listener to expect. But what is initially heard as an incongruity is invested with a new meaning by what follows, so that the momentary conflict is promptly resolved.

#### SUMMARY

The use of information-theory measures is possible whenever there is a partition and a probability distribution. The stimuli and responses of behavior theory fulfill these conditions, but the situations in which information-theory language has proved useful to psychology have been ones in which conflict is an important factor. The "uncertainty" function satisfies some of the requirements that may reasonably be laid down for a measure of "degree of conflict." But it does not satisfy them all with-

out some modification, because it depends on the relative but not the absolute strengths of competing response tendencies.

A discussion of six psychological variables that appear to depend on degree of conflict reveals several further links with information theory. The variables are emotional disturbance, reaction time, drive, curiosity, stimulus complexity, and reward.

#### REFERENCES

1. ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183-193.
2. BERLYNE, D. E. Novelty and curiosity as determinants of exploratory behavior. *Brit. J. Psychol.*, 1950, **41**, 68-80.
3. BERLYNE, D. E. Knowledge and stimulus-response psychology. *Psychol. Rev.*, 1954, **61**, 245-254.
4. BERLYNE, D. E. A theory of human curiosity. *Brit. J. Psychol.*, 1954, **45**, 180-191.
5. BERLYNE, D. E. An experimental study of human curiosity. *Brit. J. Psychol.*, 1954, **45**, 256-265.
6. BERLYNE, D. E. The arousal and satiation of perceptual curiosity in the rat. *J. comp. physiol. Psychol.*, 1955, **48**, 238-246.
7. BERLYNE, D. E. Conflict and choice time. *Brit. J. Psychol.*, 1957, **48**, 106-118.
8. BERLYNE, D. E. Conflict and information-theory variables as determinants of human perceptual curiosity. *J. exp. Psychol.*, 1957, **53**, 399-404.
9. BERLYNE, D. E. The influence of complexity and change in visual figures on orienting responses. *J. exp. Psychol.*, 1958 (in press).
10. BERLYNE, D. E., & SLATER, J. Perceptual curiosity, exploratory behavior and maze learning. *J. comp. physiol. Psychol.*, 1957, **50**, 228-232.
11. BROWN, J. S., & FARRER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465-495.
12. BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.
13. CHERRY, C. *On human communication*. New York: Wiley, 1957.
14. CRONBACH, L. J. On the non-rational application of information measures in psychology. In H. Quastler (Ed.), *In-*

- formation theory in psychology. Glencoe, Ill.: Free Press, 1955.
15. DEWEY, J. The theory of emotion. II. The significance of emotion. *Psychol. Rev.*, 1895, 2, 13-32.
  16. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, 57, 94-107.
  17. ESTES, W. K. Individual behavior in uncertain situations: an interpretation in terms of statistical association theory. In R. M. Thrall et al. (Eds.), *Decision processes*. New York: Wiley, 1954.
  18. FARBET, M. L. Suffering and the time perspective of the prisoner. *Univ. of Iowa Studies in Child Welfare*, 1944, 20, 155-227.
  19. FONBERG, E. On the manifestation of conditioned defensive reactions in stress. *Bull. Soc. Sci. et Lett. de Lodz*, 1956, 7, 1-8.
  20. FRENCH, J. R. P. Organized and unorganized groups under fear and frustration. *Univ. of Iowa Studies in Child Welfare*, 1944, 20, 231-308.
  21. GRANT, D. A., HAKE, H. W., & HORNSETH, P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, 42, 1-5.
  22. GRAVES, M. E. *Art of color and design*. (2nd ed.) New York: McGraw-Hill, 1951.
  23. GUTHRIE, E. R. *The psychology of learning*. New York: Harper, 1935.
  24. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
  25. HEBB, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, 62, 243-254.
  26. HICKS, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, 4, 11-26.
  27. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
  28. HULL, C. L. *A behavior system*. New Haven: Yale Univ. Press, 1952.
  29. HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, 45, 188-196.
  30. LOWELL, E. L. The effect of conflict on motivation. Unpublished Ph.D. thesis, Harvard Univ., 1952.
  31. LURIA, A. R. *The nature of human conflicts*. New York: Liveright, 1932.
  32. MARX, M. H., HENDERSON, R. L., & ROBERTS, C. L. Positive reinforcement of the bar-pressing response by a light stimulus following dark operant pretests with no aftereffect. *J. comp. physiol. Psychol.*, 1955, 48, 73-76.
  33. MCCLELLAND, D., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
  34. MEYER, L. B. *Emotion and meaning in music*. Chicago: Univ. of Chicago Press, 1956.
  35. MILLER, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 1956, 63, 81-97.
  36. MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.
  37. OSGOOD, C. E. Behavior theory and the social sciences. *Behav. Sci.*, 1956, 1, 167-185.
  38. PAVLOV, I. P. *Conditioned Reflexes*. Oxford: Oxford Univ. Press, 1927.
  39. PROKASY, W. F. The acquisition of observing responses in the absence of differential external reinforcement. *J. comp. physiol. Psychol.*, 1956, 49, 131-134.
  40. SAMSON, E. W. *Fundamental natural concepts of information theory*. AFCL Rep. ES079, 1951.
  41. SHANNON, C. E., & WEAVER, W. *The mathematical theory of communication*. Urbana: Univ. of Illinois Press, 1949.
  42. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
  43. SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univ. Press, 1956.
  44. STEVENS, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
  45. TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Appleton-Century, 1932.
  46. WHITING, J. W. M., & CHILD, I. L. *Child training and personality*. New Haven: Yale Univ. Press, 1953.
  47. WILSON, K. The information theory approach. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics*. *J. abnorm. soc. Psychol.*, 1954, 49 (4, Pt. 2—Suppl.), 35-49.
  48. WYCKOFF, L. B. The role of observing responses in discrimination learning: Part I. *Psychol. Rev.*, 1952, 59, 431-442.
  49. WYCKOFF, L. B. The role of observing responses in discrimination learning: Part II. Unpublished Ph.D. thesis, Univ. of Indiana, 1951.

(Received March 23, 1957)

## NEURAL MECHANISMS IN PERCEPTION<sup>1</sup>

JEROME S. BRUNER

*Harvard University*

I have been asked to offer a paper on perception, one that may be of some relevance to the manner in which the brain mediates perceptual processes. I admit to no special expertness in the neurophysiology of perception, although I have followed the growing literature with an increasing interest and a sense of excitement. There was a time, let me confess, when as a student of perception and thinking I operated on the working assumption that as far as I was concerned the nervous system was made of green cheese. This form of intellectual cheek, this know-nothing attitude, gives me little comfort nowadays. Yet many of my colleagues assure me that nothing we can find out about the nervous system can in any sense "explain" what we know about the subtle mixture of sensory, mnemonic, and inferential activity that we call perception, for, as they say, psychological theory does not reduce to neurophysiology any more than neurophysiology reduces to electromagnetic theory. I believe both of these assertions to be true. Yet, I have a lingering doubt. Indeed, I can even illustrate my doubt.

Some years ago I came to the conclusion, to simplify the matter grossly, that perceptual selectivity seemed to operate like a gating process that had the characteristic of raising the identification threshold for certain classes of stimuli and lowering them for others; and I suggested that rather than viewing such threshold differences as due to speed of

associational processes set up by different inputs, inputs might be conceived as being differentially gated out before ever reaching a stage of association. My argument was based on considerations of economy of functioning—that very likely perception had as one of its functions the task of screening out irrelevant stimuli and letting through relevant ones, an important consideration for living organisms who have a perilously narrow attention or immediate memory span. Implicitly, I was resisting a classical bit of neurophysiology as well: the doctrine about projection areas being linked up directly with association areas that enriched or "gave meaning" to sensory inputs. So I obviously was not completely independent of neurophysiological issues even at that time. During the past year, moreover, I have had occasion to read papers by Hernandez-Péon and associates (28) and by Galambos and associates (20)—both indicating that neurophysiologically, too, there seems to be evidence that screening or gating of inputs takes place as far peripherally in the auditory system as the cochlear nucleus. A cat, distracted by a mouse introduced into its visual field, ceased to show normal evoked potentials from this nucleus, far out though it was in the system. I was quite overjoyed—and in a manner not fit for somebody who claimed to believe that psychological theory must be self-sufficient—that it must not look for explanations in the black box of the nervous system.

Donald Hebb (24) provides me with an explanation of my unseemly joy. In a delightful paper in which he directs his shafts against the more unthinking type

<sup>1</sup> Presented at the 1957 meeting of the Association for Research in Nervous and Mental Disease. I am particularly indebted to Dr. Karl Pribram and Professor Walter Rosenblith for their helpful advice and criticism.



of antiphysiological bias in psychological theory he remarks, "The idea in rejecting physiology was to use only 'purely behavioral' conceptions, but some of these were actually of physiological origin and continue to exert a physiological influence on psychology" (24, p. 40). Or, as he says in a paper some years later, "... the psychologist who avoids physiological conceptions merely succeeds in avoiding modern ones, and is likely to have his thinking dominated by older ideas, vintage of 1890" (25, p. 404). And in fact, this has been the grave difficulty with physiological notions in psychology: their half life is too long, so long that their origins are forgotten and they become psychological gospel. My malaise with the associational theory of selective perception was a dissatisfaction with old neurophysiology masquerading as psychological theory. This is not to say that I wish to offer psychological theory to the contemporary neurophysiologists, but to find out, rather, what kinds of parallels and linkages might exist between what we believe to be the case in perception and what is gradually emerging concerning neurophysiological mechanisms that might mediate such processes. I might even hope that the way we as psychologists talk about perception might suggest some neurophysiological ideas to neurophysiologists!

### THREE FUNDAMENTAL PROBLEMS

There are three broad questions that one may ask about the perceptual process and these comprise, I think, most of the diverse activities that research workers concern themselves with in the study of perception; they are questions which, I think, lead most directly to neurophysiological speculation and inquiry.

The first of these is "How does perception represent the structure of the physical environment that constitutes stimulation?" Questions of this order

are familiar. How, to take an example from some work of Pfaffman (39), does salty taste get transmitted to the cortex in such a way that it can there be distinguished from quinine or acid? Or how do organisms learn to represent the probability structure of the environment, so that they may go swiftly and easily from a few cues like color, shape, and texture to the identification or inference of, say, an apple? There are far more complicated versions of this question, but these we can consider sometimes. Indeed, if there is anything to be said for the James-Lange theory of emotion, one must include here the question of how afferent impulses from the autonomic system are represented perceptually.

The second question is also simple—deceptively so, like the first one. "What accounts for the fact that perception may remain relatively constant while the physical stimuli operating at the receptor surface are varying drastically?" Here we are dealing with several seemingly different kinds of phenomena. The first is the phenomenon of constancy: the size, shape, color, and brightness constancies. In a field with distance cues available, an object moving away from the observer does not change in apparent size at anywhere near the rate required by the inverse square law that governs the size of the retinal image. Or, with field-illumination cues present, two surfaces with the same albedo appear of about the same brightness, even though one be in shadow and the other in diffused sunlight and their total luminous flux different by an order of magnitude. Moreover, for all save the very young child, objects maintain their identity though they change appearance—seen from the front or the rear, disappearing behind a screen and reappearing, going through various distortions. A range of highly variable figures are perceived and discriminated as triangles so

long as they possess certain criterial properties and regardless of other varying features, such as difference in size, texture, and so on. Indeed, the phenomena of linguistics are rife with examples of such constancy in the face of perceptual variability. Wide ranges of speech sounds are rendered constant, if they comprise the range included in a phoneme, and it is characteristic that variations within a phoneme are not noticed. Given the phonemic structure of Japanese, the naive Japanese listener does not notice the difference between "rice" and "lice," "lobster" and "robster"—the latter word having actually appeared in a Japanese-English magazine in place of the former.

The third question covers an equally wide range of problems. "What accounts for the fact that perception may vary while the stimulus input remains constant?" Here we deal with phenomena all the way from spontaneous alternation of such unstable three-dimensional figures as the Necker cube to reorganization of the perceptual field under conditions of change of set. One moment we do not understand a sentence spoken to us and the next moment we do, though it has been repeated in identical form.

#### PERCEPTUAL REPRESENTATION

When we say that perception represents the physical world of stimuli with which an organism has sensory commerce, we mean essentially that perception is varyingly predictive. That is to say, to take a banal example, if one stick appears longer than another when some distance removed from each other, they will likely show the property that one will overlap the other when placed one next the other, or that one will actually "measure" longer than the other when the technological aid of a foot rule is employed.

It was the objective of Weber (54),

then of Fechner (13), to state a general psychophysical law to describe the translation of stimulus magnitudes into subjective magnitudes. The most recent reformulation of this general law is Stevens's: "... there is a general psychophysical law relating subjective magnitude to stimulus magnitude, ... this law is simply that equal stimulus ratios produce equal subjective ratios" (46, p. 153). We shall return to this matter shortly.

Even when such simple prediction fails, the malprediction is more often than not systematic, and still contains the basis of a prediction. Add to the sticks the tails required to create the Müller-Lyer illusion and the simple foot-rule prediction breaks down; but the breakdown is systematic and, if the perceiver has the right transformation equation with which to correct his prediction, he will still come out all right. This is not to say that perceptual experience is a unique isomorph of the physical attributes of the stimulating environment, for patently it is not. There is error of a random type also operating. Partly the error can be thought of as a result of the imperfect resolving power of the sensory receptors and their related neural tracts and centers. Partly, the error can be attributed to the presence of various forms of centrifugally generated noise, or even, possibly, to the operation of centrifugal control mechanisms of the type discussed by Granit (22).

There is another form of representation which, on the face of it at least, seems of a different order. It has to do with the representation of *objects* in the environment, and such representation must obviously be learned. It consists, to use the phrase of the brilliant, and tragically dead, young Cambridge psychologist-physiologist, Kenneth Craik, in a process of model building, the construction of neural models

or "templates" to represent the redundant structures of the environment. Nowhere is the importance for perception of such representing models more evident than in the study of perception under tachistoscopic presentation. Given a brief flash of a complex object, say a picture of a human figure at 10 milliseconds in good illumination, a subject picks up a few defining cues, infers the rest, and with little difficulty and considerable confidence announces that he has seen a man, a generic man, to be sure, but a man nonetheless. If what has been presented in the tachistoscope reflects the high-probability redundancies of the environment, our subject will be right. If, however, we make up the stimulus picture with the head of a giraffe, the arms of a gibbon, the trunk of a bear, etc., then our subject will be, for practical purposes, the victim of his redundancy-matching models. We know from the brilliant observational studies of Piaget, particularly from his *Construction of Reality in the Child* (40), and from the somewhat ambiguous accounts of von Senden (44), the difficulties encountered in perceptual identification when there has been no intervening opportunity for an organism to construct such representing models for environmental objects.

Let me recount briefly some of the characteristics of perceptual representations as these have come to be known over the last half-century of relatively uninterrupted research on this topic.

1. *The categorial nature of representation.* What is most characteristic of perceptual representation is that it is categorial in nature, better described in the language of set theory and Boolean algebra than any other. An input is allocated to a class of objects and achieves its identity thereby. We see triangles, lines, apples, people. Ordering by subjective magnitude can be shown to be subservient to this first primitive,

rather conceptual placement. "Big" or "small" as a subjective judgment depends upon the summary midpoint of the class in which the object has been placed, and it is still something of a moot point how the nervous system estimates the first moment of magnitude or quality of the distribution of objects that are arrayed in a class. It is further the case that the process of achieving placement of an input is often an iterative estimation procedure.<sup>2</sup> You are trying to guess the square root of a number, say 4. Guess any number less than 4. Divide 4 by this number. If the quotient is bigger than the number you divided by, the guess is too small, and vice versa. So now take the mean of the guess and the quotient. Use this as your next guess. In a few turns, you will be very close indeed to the square root. So it seems to go, in a general way, with categorization of inputs. There is a first guess—determined by the expectancy of the organism. It is tried out and found to "fit" only approximately. The result of the attempted placement is then somehow averaged with the original guess, and on it goes. This is a gross oversimplification, but it will serve us for now. What it fails to describe is the frequent discontinuity of the process of placing, but we will come to that when we speak of the variability of perception in the face of a constant stimulus.

The iterative or, more likely, Markovian nature of successive approximation in perceptual identification also makes more understandable both the systematic nature of errors in identifica-

<sup>2</sup> I am grateful to Professor R. C. Oldfield of Oxford for bringing this possibility to my attention. The analogy and disanalogy between iterative procedure in mathematical estimation and in cognitive operations is developed in his "Some Principles of Convergence in Thinking," presented to a Symposium on Thinking held at St. John's College, Cambridge, during the summer of 1955.

tion observed in tachistoscopic perception and the abrupt manner in which a long series of increasing errors sometimes gets corrected. If, for example, a subject is presented successively at rapid exposure with a stimulus array containing an incongruous element—say, a picture of a discus thrower, wound for the throw but holding a bass viol in his free hand—there will be a series of iteration-like attempts to deal with the incongruity soon after the perceiver is able to report that he perceives an athlete in action. It is a “shadow across the front of the athlete” or “a track official kneeling” or a “sprinter going past on the track,” and these hypotheses become elaborated on successive exposures: inhomogeneities are seen as the track official's hat or as a sweat suit on the passing sprinter. Finally, the elaboration reaches a point where its lack of fit become obvious and there is a notable disorganization of perception, until a new series of approximations begins (cf. Bruner and Postman, 7).

Perceptual identification or placement, then, can be likened to the determination of the fit between a model and some sample that is being matched to it. We know precious little about the nature of the models that are constructed neurally to represent the attributive and probabilistic structure of the individual's environment, although various models have been proposed, depending in the main upon the development of reverberatory circuits and synaptic boutons, as in Hebb's conception of cell assemblies and phase sequences (23), or upon reduplication patterns propagated across the entire cortex, as in Lashley's ingenious conception of memory traces, concerning the failure of whose location he writes so provocatively (32). But it is not about these structures that I wish to speculate, for there has been enough speculation on this subject in ratio to the meager findings available. Rather,

it is with the process of matching or placing of inputs that I wish to concern myself briefly.

We know that rhythmic activity from the surface of the brain is related to attention, and that alpha rhythm is disrupted and replaced by more aperiodic rhythms when attention is directed to a stimulus. Let me put this finding together with two others. The first relates to the rhythm recorded from the olfactory bulb of the rabbit by Adrian (2), a rhythm of high frequency of 80–100 per second that continues so long as the animal is breathing regularly. If an odorous substance be introduced or the animal begins to sniff, “the rhythm disappears or gives place to the slower rhythm which occurs at each aspiration and is due to the olfactory discharge. Within a short time, 20 seconds or less, the rapid rhythm returns although the olfactory stimulation continues . . . but a change in the stimulus may suppress it again.” Adrian then goes on to say,

The organization of the olfactory pathways may be quite different from those for other forms of sensation, but it is natural to ask whether the rhythms found in the cerebral cortex may not operate in the same way, whether their function or part of it is not to block the passage of further afferent discharges when the information has served its purpose. The reticular formation might well be the decisive factor in the direction of attention, i.e., in the suppression of the rhythm in one region, for the signal arriving by the direct route to the cortex reinforced by one from the [reticular] formation might disrupt the rhythm and gain a clear path, whereas a signal not so reinforced would be unable to break the barrier (2, p. 243).

It should be noted immediately, however, that too much arousal reaction may have the effect of interfering with the signal arriving at the cortex. Psychophysical experiments show, on the whole, that too alert an observer pays a price in a raising of sensory thresholds, and that a certain amount of relaxation in the observer yields the lowest thresholds.

Yet, this special point aside, Adrian's observation fits well with what we know of the operation of attention.

A recent paper by Sharpless and Jasper (45) provides parallel and more elaborated evidence on the role of arousal reactions—evidence on the habituation of such reactions. These writers point out that it has been known for twenty years that with direct recording from the cortex of unanesthetized animals one may observe habituation to such ordinarily attention-compelling stimuli as sharp whistles and claps. The authors set themselves the task of analyzing the locus and the specificity of such habituation effects, recording from the cortex and from subcortical areas in cats with chronically implanted electrodes. Stimuli were repeatedly presented to normal sleeping cats whose electrocorticograms showed a clear sleep pattern. With habituation to a given stimulus, the duration of activation declined and its latency increased strikingly, sometimes showing a 20-second period between stimulation and activation onset. Eventually, complete habituation would occur. These patterns and the habituation effects were recorded from the cortex as well as from the mesial thalamus, the posterior hypothalamus, and the midbrain reticular formation. With respect to the specificity of habituation, a critical point in our discussion,

If either the intensity or duration of a stimulus to which the animal had become completely habituated was increased, it became capable of arousing the animal. More interesting, however, was the specificity of the habituation process to the quality of the repeated stimulus. Thus, if the animal had become completely habituated to a repeated tone, it could still be aroused by a light touch, a change in illumination, or by a tone of a different frequency (45, p. 662).

While the authors report some generalization of habituation to tones differing somewhat in frequency, they remark that

it is variable. They note also that "habituation was occasionally specific not only to the pitch of a sound, but also to the arrangement of tonal elements within the sound—its pattern." If a cat had been completely habituated to a tone falling in pitch from 5,000 to 200 c.p.s. in 4 sec., presentation of the stimulus pattern in reverse—rising from 200 to 5,000 in the same time—would produce activation. Destruction of the auditory cortex did not alter the pattern of habituation observed in intact animals nor its specificity—save for habituation to pattern, which was not found in such operated cats. Subcortical lesions, however, seriously disrupted specificity of habituation to tonal pitch, notably subcortical lesions that prevented direct passage of signals to the cortex but left intact the collateral pathways to the reticular system through which auditory signals could get to the hemispheres.

In sum, then, the evidence of Sharpless and Jasper seems to indicate that the arousal pattern that may subserve the selectivity of attention is capable of considerable selectivity of habituation to environmental stimulation. If habituation has this degree of specificity, it seems not unlikely that the obverse process, selective attention built up through training, might operate by a parallel subcortical mechanism—selectively reinforcing not only "non-habitual" stimuli but also stimuli to which, through training, special significance had become attached. Such reinforcement may be essential for the stimulus input to reach a level of activation where a "fit" can be made between an input and a "model."

The test of this is in principle not difficult to make although it is probably very difficult technically—at least for the while. If it can be shown that there is an onset in arousal impulses from the reticular system that occurs while an observer is trying to "figure out" the nature of an input, and a dimi-



nution following the identification of the stimulating object, then it would seem more plausible still that "effort after meaning," to use Bartlett's phrase, is dependent for its fruition upon reinforcing stimulation from the reticular system—either to clear the track for it, or to make possible the process of matching.

2. *The summary nature of perception.* The act of perceiving almost always is accompanied by a loss of information. That is to say, if we think of the perceptual process as an input-output system, we find that no matter how we measure the nature of the input—either by physical measurement or by a consensus of observers—the limited attention span of the human organism (and presumably lower organisms as well) imposes loss in the transmission of information, even assuming that internally generated noise could be eliminated. This seems to me to be intuitively obvious, though it ends one in much complication trying to measure the nature or extent of the loss. The only known way of combatting loss in transmission is by a process of what Miller (37) has recently described as "recoding of input." If it is the case, and the evidence speaks rather overwhelmingly for it, that attention or immediate memory span is for about seven units plus or minus two, the recoding process consists of regrouping the input such that each of the seven places, so to speak, is filled with more valuable coin. A good example is the phenomenon of segmentation in language perception—perception either of speech or of written language. Presented with an array of letters, we do not try to take in the individual letters, but to recognize words. While we have a span of from eight to ten random letters, the span for letters when grouped into words is much greater. So too with the speech flow and its segmentation into words on the basis of phonemes. But no recoding is so perfect that it

makes it possible to scan a great slice of the universe and recode it without information loss, for most of the forms of recoding we employ fall considerably short of such comprehensiveness, so that wherever we turn there is information loss.

Now may I quote a debate reproduced in the volume *Brain Mechanisms and Consciousness*, the symposium on the reticular system held in 1953, a debate between Bremer and Fessard. Bremer is criticizing the idea that integration of perception occurs at a locus; the particular locus he is decrying is the reticular formation. "I would certainly not have chosen the reticular formation for that location on account of the fact that information is lost there by reason of neuron convergence and occlusion. Conscious integration excludes dilution and loss of information" (5, p. 245). Fessard defends his conception of locus, and particularly the locus of the reticular formation, in part in these words, with which I thoroughly agree without taking sides on either the notion of a specific locus or the role of the reticular system:

Dr. Bremer has judiciously pointed to its main difficulty, that is, to the loss of information that would result from the simultaneous arrival of impulses from different origins upon the same neurone, as we now know that this happens in the reticular formation. I admit that this difficulty is not easily disposed of, but I do not agree with Dr. Bremer when he says that integration cannot go with loss of information: psychologists would certainly support the idea that integration is only obtained at the cost of big losses of detail, as when a perceptive Gestalt is apprehended as a whole, its components being totally ignored (14, p. 248).

I should like to propose, as an hypothesis, that both Bremer and Fessard are correct—the former in denying locus to integration, the latter in assuming that loss through convergence and occlusion is a feature of afferent transmission, just as loss of information is a fea-

ture of perceiving generally. Two lines of argument and evidence seem to recommend themselves here. The first is Lashley's argument about summation in the afferent system; the second is the point of view put forward by Rosenblith (42) in discussing the bioelectric fate of a click from the round window to the auditory receiving area.

Lashley, as you know, finds the summative characteristic of neural activity a necessary parallel to the nature of conscious experience itself.

A point of fundamental importance for a theory of the neural basis of perception is that there is never awareness of the integrative activity of the brain while it is in progress. The perceived items are always the product of preceding and complex integrative processes. Visual distance is a good illustration of this. Things are seen as near or far, yet this distance is actually determined by a number of variables, binocular parallax, estimates of relative size, texture, etc. which are not separately perceived but are only revealed by experimental isolation. . . . So in every case, that of which we are aware is an organized structure; the organizing is never experienced (33, p. 424).

Lashley then goes on to consider neural summation.

Consideration of the role of summation in neural activity reveals a similar characteristic. Even at spinal levels summation of excitations is essential for reaction. Summation combines a multiplicity of excitations into a single effective unit which the resultant reaction cannot analyze into separate elements. So also we must conceive of higher level integration as a sequence of excitations, individually incapable of exciting further activity, but continually summing to arouse the succeeding neural patterns (33, p. 424).

It is perfectly apparent that given such a model of summation—and I shall assume that it inevitably involves stages of convergence and occlusion—it is not necessary to suppose that integration has a particular locus. It is the entire range of summative activity, from the brain stem and lower on up to the cortex, that enters into integrative action. I will

make bold to propose that an understanding of the *nature* of this summative activity is what will provide the key to an adequate understanding of how it is that perception loses information both systematically and randomly, and thereby manages to achieve integration and summary representation.

Let me turn briefly to one proposal concerning the nature of summative activity, to a model inherent in Rosenblith's conception of click transmission in the auditory system.

We are inclined to interpret the electrical responses that our wire electrodes record as summated activity from a sample of a population of neural elements at a given station of the auditory nervous system. The number of neural elements that contribute to a response depends upon the intensity of the stimulus that is used. As has been shown [by McGill and Rosenblith (36)], the amplitude of an evoked response can—within certain limits—be interpreted as an estimate of the probability of response of the neural population under investigation (42, p. 5).

Rosenblith is concerned in this passage principally with the problem of amplitude of response, and it may well be that the statistical operation of the auditory and other systems works by a simple system of unbiased sampling-summation where amplitude is concerned. But there is also good evidence that higher-level units respond to biases in the sample of lower-level units that summate to activate them. The best evidence of this is provided by close analysis of the systematic patterns of discharge that occur at lower levels in response to known physical properties of stimulation. Again audition serves us well. Consider the second-order neurones in this system, as studied with microelectrodes by Galambos and Davis (18, 19). As Davis summarizes it:

Each second-order neuron is "tuned" to a particular frequency. Less acoustic energy is required at this frequency than at any other to set up impulses. Near threshold, the tun-

ing is very sharp, but the neuron responds to a broader and broader band of frequencies as the intensity is increased. . . . Even for fairly strong tones the tuning of each neuron to a particular frequency is still apparent (9, pp. 1128-1129).

This is the situation at the second-order neurones, after transmission across a single synapse. Unfortunately, the data available on the pattern of firing in the spiraling single neurones of the short eighth nerve (47) do not provide a simple picture of how firing patterns in the first neurone signal across the synapse to the second neurones of the auditory system, nor do they indicate whether there is a summation pattern operative there. When we tune the organism to listen to pitch or to loudness or to density of a tone, or to some particular pitch or loudness or density, it may well be that the nature of the summation required for firing a higher-order unit changes as far down the pathway as the first synapse, and that the resultant final message to the auditory receiving area reflects these changes in nature of summation. We must beware of being too impressed by the elegant and relatively invariant cortical mapping of auditory frequencies found for the anesthetized dog's brain by Tunturi (50, 51). Anesthesia plays funny tricks and, as Békésy (4) has suggested, "for every level of anesthesia we will probably have a hearing theory."

What I would like to suggest—and I have no neurophysiological warrant for doing so but only a hunch based on psychological observation—is that the form of biased sampling of lower-order units that determines the summated activation of higher-order units is very likely controlled by centrifugal control mechanisms that serve to program lower-order stations in the afferent system. These may be thought of as orders for types of summation that can effectively trip off higher order neurones on which lower

order neurones converge, and it may well be that such changes in nature of summation account for the way in which the summary nature of perceiving changes as a function of change of set in the organism. Adrian puts the matter metaphorically:

. . . in normal waking life the picture that is presented to consciousness has many details left out. First one element of the pattern claims our attention, then another. At some stage the complete report from the sense organs must be subjected to an editing which emphasizes the important items and sets the unimportant aside. There is clearly some such editing at work in most of the sensory pathways. . . . It might be expected wherever the signals pass through a sheet of closely connected neurones (2, p. 239).

He then goes on to suggest two possibilities as to where this editing might occur:

One is that the afferent signals from, e.g., the ear may be summarized on their way up the auditory pathway, but that the content of the message is allowed to reach the cortex whatever may be going on in the other sensory pathways. This means that at any moment the cortex would have at its disposal the complete pattern of the environment provided by the sense organs, but that there is some controlling mechanism (of which the reticular formation may be a part) to decide that the auditory elements of the pattern should have priority and to make the auditory region of the cortex specially receptive and the other areas less receptive. The other possibility is that the controlling mechanism operates at an earlier stage and that we attend to the sound and are unaware of contacts and pressures because the auditory signals are allowed to reach the cortex and the tactile signals are not (2, pp. 239-240).

Is it not conceivable that successive "synaptic editing" could occur, controlled by centrally induced programming of summation characteristics, and that there could also be the kind of "boosting" operation discussed earlier wherein a system like the reticular formation reinforces the edited pattern of stimulation that reaches the brain?

To sum up, then, we have proposed

that information loss in perception can be viewed as the resultant of successive summation at stations along the afferent pathways, that this summation is likely statistical in its characteristics, that the sampling of lower-order units that summate to fire higher order units is biased by a gating or programming of way stations along the pathway controlled by more central mechanisms, and that the resultant in perception is a biased summary of the external stimulating environment. The biasing is reflected in the systematic errors of perception and in the forms of selectivity that make it possible for perception to vary in the face of a constant pattern of external stimulation.

3. *The autogenic properties of representation.* If it can be said that the representation of the external stimulus environment is categorial and summary in nature, we must add to these properties that representations, once accomplished, appear to generate processes of their own. Here we must consider briefly some of the properties of the models that organisms construct to represent their environment and its objects. It was suggested some years ago by Craik (8) that neural models, once established—whether in the form of traces, schemata, cell assemblies, or whatever short-hand terms one uses—have the property of making imagery possible in the absence of stimulation. A recent work by Mandler (35) on the "learning of imagery" suggests that some such function is served by learned structures. His experimental procedure is very simple. The subject has the task of learning the correct pattern amongst a set of spatially arrayed levers on a board. The order of correct levers is determined at random. In a given number of trials, the subject masters the correct order in which switches are to be thrown; he is able to perform at a criterion of 100% correct responding. The trials are now continued and

overlearning carried many trials beyond the number needed for mastery of the task. After a certain amount of overlearning, subjects begin to report not only that they can choose the right path but that they have a spatial image of the order, a spatial image that they can keep in mind even when the lever board is not before them. It is tempting to speculate what this means. When neural integration has progressed to a certain level, the organization achieved appears to be able to generate mental content of its own. I do not pretend to understand what kinds of neural mechanisms are at work here, but I commend the problem to you as perhaps worth pursuit by the use of electroencephalographic methods. When learning has progressed to the stage of overlearning and imagery production, are there detectable changes in the pattern of EEG to be recorded? Methods for integrating and analyzing complex wave forms now widely in use in work on EEG might cast some first light on this problem that would permit more enlightened next steps.

There are other autogenic properties of what I prefer to call operative cognitive structures that are worth a few words of description on the chance that they might make contact with work now going on in neurophysiology. The first of these is a *normalizing tendency* exerted by cognitive models on the percepts they generate when stimulus input occurs. Let me first give an example and then explain what I mean by this process, usually called assimilation. For each category or model constructed to represent objects or events in the environment, there appears to be a set of average property values characterizing the typical instance of the class. It may be likened to the adaptation level in psychophysical judgment about which Helson (26) and others have been concerned. It is a reference point in the class which, in judgments of brightness

or weight, is best estimated by computing a weighted geometric mean of instances previously encountered. Subjective magnitudes greater than this midpoint are described as "heavy" or "bright," smaller as "light" or "dark." With respect to more complex objects, say apples, there are a set of such specifications of a typical apple—a modal color, shape, size, texture, etc. In the event of masking noise or of ambiguity in the nature of a stimulus input, one finds normalization of the percept taking place by assimilation of aberrant or ambiguous attributes to the modal values of the category where the input has been coded. Or, in the case of inattention to particular attributes, modal values of the category will be substituted for what has not been noticed. Such normalization or completion is one of the principal sources of error in perception—error in the sense of failure of precise representation. An average is substituted for a specific value. The normalizing process continues, moreover, in the process of forgetting—forgetting being more often describable as regression away from specific characteristics of stimulus event in the direction of the modal characteristics of the category into which the stimulus event has been placed.

I readily recognize that neurophysiological research has not proceeded far enough to have many sound conjectures about the kinds of structures that can produce the types of autogenic activities just described. Yet it is well to bear such characteristics in mind in the process of setting forth hypotheses about the structures. Looking at the field from the outside, I have the impression that the neural mechanisms proposed, even in a half-playful spirit, tend to be built for the purpose of handling a single psychological phenomenon. There are too many plausible neural models of the one-phenomenon type. It is only when one

requires of them that they accomplish several forms of psychological resultant that the range of possibilities narrows realistically.

#### PERCEPTUAL INVARIANCE AND VARIABLE STIMULATION

Let me begin with Fletcher's well-known study of speech perception (16), a striking example of what Lashley has for a long time called "the equivalence problem." A sample of speech is presented, treated in such a way that all frequencies *below* a certain cutoff are filtered out. Now the same speech sample is presented again, this time all frequencies *above* the cutoff being filtered out. In both cases, the speech pattern is 100% intelligible to listeners. Frequency, we then say, is not a criterial attribute of speech; one band of frequencies will do as well as another. And so it will go with most forms of perceptual identification, given, as we have noted, its categorial nature. The task in any given study of perceptual equivalence is to determine by reduction methods what particular stimulus attributes are criterial for identification of an object as belonging to a certain class or, more simply, as being a certain thing. Formally speaking, the determination of the criterial attributes of an equivalence class is the same whether one is at the perceptual level or at the level of providing the defining properties of such an abstract concept as force or momentum.<sup>3</sup> My concern here is to consider what is involved in the more perceptual forms of such activity.

Virtually every writer who has tried his hand at a neural theory of perception has been mindful of the problem; it is not overlooked. Adrian perhaps puts the matter most succinctly:

<sup>3</sup> For the reader interested in the continuities in all such categorizing I would suggest the work of Bruner, Goodnow, and Austin on categorizing and thinking (6).



We can recognize a tune whether the pitch is high or low, and we can recognize a triangle or a letter, whatever its size or whatever the position of the image on the retina. Now this must mean that there is no need for particular sensory endings or particular pathways to the cortex to be excited; for the recognition of the pattern all that seems to matter is that the excited regions should have some general relations to one another. There must be something corresponding to the three angles of the triangle, but there is no need for each of them to be represented at a particular point in the receiving area of the cortex (1, pp. 81-82).

There has been much bitter ink spilled on the question of equivalence, notably amongst psychologists who tend to take one of two basic positions. One of them, stemming I think from neurological theories of a half-century ago, is that equivalence occurs by the formation of a generalization gradient, a kind of irradiation from a particular stimulus value. The model is familiar from Pavlov's work on generalization, and it is this model that has dominated much thinking in American psychology. The organism learns to respond to a given value of a stimulus or to a given object plus a range around it. Lashley (33) has been particularly forthright in attacking this view, noting that the model probably holds only for instances of pseudogeneralization where there is failure to discriminate the original stimulus from one presented later. There is much that is logically awkward about the simple generalization notion that various writers have tried to patch up in one way or another—some quite plausibly. It does not explain how certain properties come to be the defining properties of a class and others irrelevant. To handle this, a reinforcement theory is invoked, and the defining properties are said to be determined by the composite attributes that are always present when reinforcement occurs, the remainder averaging out in strength because of their random relation to reinforcement. But, as Woodworth (55) has wisely noted, what can we take as

reinforcement in perception? In most theories it is taken to be something that reduces the need of an organism, related to attaining a goal. To extend this to the question of learning the properties of the class "triangle" is somewhat difficult. In any case, since the controversy does not concern us directly, we can leave the matter with the statement that the neural model of a class assumed to be constructed in this way is a bit like a composite photograph in which the tried and true elements remain, and the others are blurred out. Such a model of integration, suitable though it may be for describing how it is that the three angles emerge as the diacritica of triangles while size and texture and relative length of sides are blurred out, fails to deal with the essentially transformable properties of groupings once they have been formed; nor does it show how it is that in one moment of insight one comes to group all plane figures as resolvable into combinations of triangular forms and treats all plane figures as members of a new class of simple and composite triangles.

It seems to me that the problem of equivalence, leaving aside now how the primitive equivalence category is formed, is essentially the problem of how an organism processes a stimulus ensemble, what features of input it is responding to. It is only with respect to this program for processing environmental stimulation that it is possible to talk about an adequate stimulus or a just noticeable difference in stimulation. Almost invariably, as both Adrian (1) and Lashley (33) have suggested, stimulus information is processed in terms of *relationships* either in a spatial array or in a temporal array, and it seems highly likely that spatial relations are translated into temporal ones in the nervous system—given the requirement of time for scanning a spatially extended field. What this suggests, then, is that the

equivalence of stimulus events is a function of certain invariances in relationship in a temporally extended neural excitation; certain patterns of change in excitation define equivalence.

There is very suggestive evidence concerning the importance of temporal change as providing the pattern on which equivalence is based. A famous passage of Exner's (12), where he compares the scanning movements of the copepod *Copilia* with the scanning movements of the eye, comes to mind, "Wir tasten mit unserem Blick des Seefeld ab," and indeed it is this "fingering over" the visual field that suggests the importance of temporal integration. Thorpe (48), in his recent and most excellent book on animal behavior, gives various examples of the importance of temporally extended scanning of a stimulus array in animal discrimination. Dethier (10), for example, has shown that the caterpillars of *Lepidoptera* can respond to form and form equivalence in spite of the optically inadequate stemmata placed on either side of the head, optically insufficient in resolving power for representation of form. Yet form can be perceived by these organisms provided it is possible for them to move their heads around in relation to the object to be discriminated, by the familiar waving or "inching" movements of the anterior part of the body of these caterpillars. Some sort of scanning process, extended over time, results in a temporal summation and/or integration that makes form discrimination possible. So, too, in the work of Ulflyott (52) on klinokinesis and klinotaxis in the flatworm *Dendrocoelum*. Again, a primitive optic apparatus is involved, supported either by the animal's locomoting by a convoluted path with a relatively constant rate of change in direction (klinokinesis), or by proceeding in a relatively straight line with rather constant waving of the an-

terior section of the body where the sense receptor is found (klinotaxis).

Given the nature of adaptation of sense organs, it is not unlikely that equivalence coding almost requires the use of patterns of change in excitation over time. Not only do we have the facts of adaptation with respect to intensity of stimulation in vision, somesthesia, etc., but there is now accumulating evidence from the work of Riggs (41) and of Ditchburn (11) that if change of pattern on a sensory surface does not occur at a quite rapid rate, perception of form and detail adapts out. Ditchburn, using an ingenious mirror attachment on the cornea, casts a simple form image on to it, collects the image, and directs the beam back into the eye. The arrangement keeps the cast image at a constant point on the retina, resting on this point in spite of eye tremor. Within six seconds, the simple image, a fine-lined reticule, disappears. The moment one jars the apparatus, even slightly, the image changes position however slightly on the retina and reappears. These results leave little question that change over time is needed even for the maintenance of a simple image.

I should like to suggest that equivalence of an array of stimuli is essentially a centrally determined program concerning the acceptance limits of temporally extended change patterns. Where and how these patterns are acted upon and coded for equivalence, I have no clue. Adrian and Moruzzi (3) show suggestive evidence that the electrical pattern in the somatic cortex is much alike when the same place on the foot is touched, so that it is possible that the cortex has enough information to do all the job of coding for equivalence in terms of some program of temporal integration.

Let me say, finally, in support of this general view, that most of the findings in the constancies seem to support the

view that apparent brightness, color, speed, and form—all of them exhibiting notable tendencies toward high perceptual invariance in the presence of marked change in the physical stimulus—turn out to be explicable in terms of constant relationships existing in the stimulus ensemble. Wallach (53) has shown that, within wide limits, apparent brightness is a function of ratios of stimulation in the visual field. A central field surrounded by an annulus with a ratio of 1:10 in luminous flux appears to be the same brightness as a central field that bears a relation of 10:100 to its annulus. Two objects are seen as traveling at the same speed if their angular displacement bears a similar relationship to their apparent distance from the eye—a more complicated form of ratio. And as Gibson (21) has shown, apparent distance is a function of the change in texture of the field intervening between the object and the eye. He distinguishes four temporal or spatial orders: Let *l* and *d* indicate light and dark. Then the order *lllll* or *dddd* will be seen as a real or film color; *lllddd* or *dddlll* as a line or the edge of a figure; *ldldldld* as a surface; and *llldddllldddlllddd* as a gradient of texture indicating distance, equal distances presumably being based on like gradients of texture, all other things being equal.

Undoubtedly, some of the equivalence coding in the perceptual constancies is innate, some learned. When we come to identity constancies, treating as equivalent objects that have been altered drastically in all respects save their defining attributes, we are very likely dealing with learned behavior exclusively. And here the "rules of equivalence" that are learned doubtless reflect the adjustment requirements of the organisms. We know from the work of Thouless (49), for example, that mechanics must and do learn to *depend* on size-constancy effects in judging the sizes of nuts and

bolts and the like, whether the objects are held in hand or are seen on the bench at some remove; and that painters learn to *overcome* brightness and color constancy effects to get on with their jobs. Klüver's brilliant monograph on equivalence of stimuli (30) indicates the degree to which equivalence grouping is generally subject to learning. But perhaps the matter will be better elucidated if we turn now to the problem of perceptual variation in the face of constant stimulation.

#### PERCEPTUAL VARIABILITY AND STIMULUS CONSTANCY

Let me begin by telling you of a game my colleague George Miller and I played at lunch one day. The question was whether one could construct a sentence pun rather than a word pun. Our two successes serve to underline the nature of selective organization of a flow of information. The two sentence puns are, first,

The good candy came any way.  
The good can decay many ways.

and

One's own leeway is to earn ever.  
One's only way is to err never.

Spoken rapidly, each speech flow in a pair is virtually identical from a phonemic point of view. In the first pair, a final *s* provides the cue for retroactive segmentation of the flow into lexemes or words. In the second pair, there is no such cue provided. The perception of speech is full of such examples of differential ways of organizing a temporal flow of stimulation. Since the pioneering work of de Saussure (43) on the isolation of the phenome, and with the development of Jakobson and Halle's method (29) of decomposing the phoneme into a set of distinctive features, we know that the process of understanding speech involves a highly selective

method of isolating certain ranges of speech sounds, treating these quite arbitrarily as equivalent, and then using these as the diacritica by which words may be distinguished. There is often ambiguity in the process of segmentation, but this is rarely serious, for context almost always settles the issue—although the low-comedy deaf clown of classic vaudeville testifies to how close a thing such settlement can be.

It is a commonplace of psychological research that the organization of a complex perception can be varied by varying the set of the person—by varying the thing for which he is looking. Or, to use another common form of discourse, what we take in depends upon how our attention is directed.

What are the mechanisms that determine whether or not a given stimulus is perceived? What determines its dominance over other stimuli? Very intense stimuli generally dominate, and certain sensory modalities seem generally to be prepotent. This is particularly true of pain, perhaps of movement in the peripheral retina, of high pitched sounds, of some odors. In birds and higher primates vision is prepotent but probably not in other mammalian orders, though the evidence is uncertain. However, for man, and probably for most animals, the characteristics of the stimulus are generally of secondary importance in determining what is perceived. Internal factors, characterized as voluntary attention, play the major role in selecting stimuli for perception. Sometimes such attention involves chiefly giving prepotence to a sensory modality, as when one looks at a picture or listens to music. Even in such cases, however, there is a further selection of the items perceived, in terms of organization that is somehow inherent in the neural processes (33, pp. 426-427).

I think Lashley's summary of the matter admirable. What manner of neural processes seem to be involved?

Adrian proposes a first step toward an answer.

The operations of the brain seem to be related to particular fields of sensory information which vary from moment to moment with the shifts of our attention. The signals from

the sense organs must be treated differently when we attend to them and when we do not, and if we could decide how and where the divergence arises, we should be nearer to understanding how the level of consciousness is reached. The question [is] whether the afferent messages that evoke sensations are allowed at all times to reach the cerebral cortex or are sometimes blocked at a lower level. Clearly we can reduce the inflow from the sense organs as we do by closing the eyes and relaxing the muscles when we wish to sleep and it is quite probable that the sensitivity of some of the sense organs can be directly influenced by the central nervous system. But even in deep sleep or coma there is no reason to believe that sensory messages no longer reach the central nervous system. At some stage therefore on their passage to consciousness the messages meet with barriers that are sometimes open and sometimes closed. Where are these barriers, in the cortex, the brainstem, or elsewhere? (2, pp. 238-239).

I would like to suggest as a first hypothesis that the barriers are everywhere. The recent experiments of Galambos *et al.* (20) and Hernandez-Péon *et al.* (28) suggest that in the auditory system the barriers may exist as far peripherally as the second synapse at the cochlear nucleus. In the first of these studies it was shown that evoked potentials in the cochlear nucleus produced by click stimulation adapted out with repetition. The click was then paired with an electric shock delivered across the chests of the experimental cats, and conditioning established; following this, the conditioned reaction was extinguished and then re-established, etc. While the click was a signal for shock, there was greater electrical discharge from the cochlear nucleus as well as from higher centers—the caudate nucleus and the auditory cortex. When extinction occurred, when the animals did not show the crouching-snarling response to the click, there was marked reduction of activity in all three loci. To check whether the motor activity associated with the conditioned response was responsible for these changes in evoked potentials, Ga-

lambos and his colleagues checked their findings on cats in whom a complete muscular paralysis had been induced. No difference in the general findings was produced by this precautionary control.

The experiments of Hernandez-Péon *et al.* were of the same order, save that the attentional factor was varied in a somewhat different way. Again, activity was recorded at the cochlear nucleus, involving either second- or third-order neurones, and cats were exposed to pure tones at intensities comfortable to the human ear. Electrical activity was studied with three forms of distraction outside the auditory sphere: while exposed to the tone, cats were shown two mice in a bell jar, were given fish odors to smell, or were shocked on the forepaw; i.e., visual, auditory, and somatic distractors were employed. Electrical activity at the cochlear nucleus was markedly reduced under all three conditions in comparison with control responses. The authors propose that the suppression of the response in the cochlear nucleus is probably effected by inhibitory impulses from the midbrain reticular system.

The question arises whether the inhibition and facilitation effects found in these two studies are selective within a modality, whether there is selectivity for particular tones or other stimuli. The evidence here is still sketchy. Hernandez-Péon and Scherrer (27) have reported, however, that selective adaptation or habituation effects, known to occur in the midbrain reticular formation, can also be shown to operate at the cochlear nucleus. Prolonged repetition of a tone of a given frequency produces habituation in this nucleus, but if the frequency of the stimulating tone is changed, the electrical activity in the nucleus increases to normal levels. What remains to be done is to repeat the Galambos experiments with changes in the characteristic of the stimulating tone.

If the conditioned stimulus is a tone of 3000 c.p.s., will a tone of 5000 c.p.s. produce less electrical discharge in the cochlear nucleus, the caudate, and the auditory cortex?

Space does not permit a proper appraisal of the evidence for the assertion that selective barriers may operate in altering perception at virtually any level of the nervous system. Granit (22) presents evidence that suggests such a widespread locus of gating, inhibition, and facilitation. The work of Leksell (34) and of Kuffler and Hunt (31) on the control of afferent impulses from muscle spindles by the small diameter fibers of the ventral root of the spinal cord suggests that gating and control of impulses can occur right at the sense receptor. I do not know whether those of you who are expert in such matters would urge that, if there is such widespread monitoring of incoming impulses, it can all be traced back to the action of structures in what Penfield (38) calls the "centrencephalic system." I know that Professor Lashley's is a dissenting voice (33) to what appears to be a growing chorus of assent. Pribram<sup>4</sup> is also among those who would look for a broader locus, including a considerable role for what he speaks of as the posterior intrinsic sector of the dorsolateroparietal convexity of the cortex. He distinguishes in perception between "detection" and "identification," the former being a judgment of presence-absence of a stimulus and the latter being its placement or recognition. Destruction of the posterior intrinsic sector in monkeys appears to destroy their capacity for choice discrimination problems while leaving seemingly intact their capacity to respond to presence and absence of a stimulus. But it is interesting to note that Pribram describes the posterior intrinsic sector as receiving its extratel-

<sup>4</sup>Pribram K. The neurology of thinking. (In preparation.)



encephalic input exclusively from the thalamus, and it may well be that here again we are dealing with a central alerting or reinforcing station in the manner of Penfield's "centrencephalic system."<sup>5</sup>

But these are surely not questions of detail to be settled by a psychologist not properly cognizant of the anatomical and electrophysiological subtleties involved in locating where gating mechanisms are operative. My only reaction is that what is psychologically characteristic of selectivity on the afferent side is that one finds it at all levels of functioning—in the reception of information, in the selective losses of memory, and in the transformations of information involved in the thought processes. It seems highly unlikely, given this widespread range of selective loss and transformation, that there is one "seat" of this function in the nervous system, although it does not rule out the possibility of a centralized coordinating sector. Indeed, Galambos (17) has found not only that gating extends down to the first neuron of the auditory nerve, but that one can create gating at this level by artificial stimulation of the olivo-cochlear bundle, and he also shows that section of this bundle prevents such centrifugal gating. Since this has been demonstrated, one wonders

whether we are not on the verge of a new era in neurophysiological research where the old model of the reflex arc is once and for all going to be dethroned. The model that is emerging is certainly one that accords more with the known variability of perception under conditions of constant stimulation.

Finally, let me say a few words about the nature of variations in perceptual organization as these occur in the presence of a constant or steady stimulus. In the preceding section, mention was made of a set of rules governing the manner in which the relations within a stimulus ensemble might be used for coding the input, and that so long as inputs showed the prescribed relationships in the patterns of excitation they elicited, they would be coded equivalently. The complementary point can be made about variability in perception with constant input. Given a change in the rules for the utilization of stimulus relationships, the resulting percept can change even though the stimulus remains the same. Pribram,<sup>6</sup> in writing on the neurology of thinking, remarks that the transforms that govern the utilization of features of an excitation—what we speak of as "rules"—are determined by factors other than the nature of the immediately operative stimulus input. The question of what kinds of events bring about such alterations in rules for gating afferent stimulation is still an open one, and it is still a matter of debate whether there is some central activating system that is always and necessarily involved in such alterations.

#### A BRIEF CONCLUSION

The model of perception we have explored in these pages is one that is a drastic departure from the conventional stimulus-response, associational, or reflex-arc model that is a legacy to psy-

<sup>5</sup> Since the above discussion was written, the writer has been informed by F. Bremer of Brussels of an experiment carried out in his laboratory by M. Jouvét and J. E. Desmedt in which the electrical activity of the cochlear nucleus in response to a click is inhibited by concurrent electrical stimulation of the lateral part of the anterior mesencephalic tegmentum. The inhibitory control exercised by this mesencephalic locus does not appear to be mediated by the tract joining the superior olive and the cochlear nucleus. In any case it is becoming increasingly apparent that there are likely to be many centrifugal routes for the control, monitoring, and regulation of afferent input. A brief report of the work of Jouvét and Desmedt is to be found in *Comptes rendus des seances de l'Academie des Sciences (Belgian)*, 243, 1916-1917, seance de 5 decembre 1956.

<sup>6</sup> See footnote 4.

chology from the neurophysiology of a past generation. It is a model in which, to use the language of Fessard (15), there is a de-emphasis on transmission of impulses across synaptic segments and a shift in emphasis to integration and autogenic activity—a model including complex networks with the capacity to hold up and to alter the characteristics of impulses transmitted to them, and with the capacity to initiate activity that is transmitted elsewhere to affect control of afferent impulses traveling to the cortex and efferent impulses traveling away from it. It is a system that, to put it figuratively, can determine within considerable limits the nature of the effective excitation which results when a change in physical energy impinges upon a sense receptor. The tracing and manipulation of efferent fibers carrying impulses to synapses along receptor pathways and to sensory receptors themselves indicate that the neural model we shall be using is one in which centrally induced control of sense data will play an increasing role. If the neurophysiology of a generation ago forced psychology into a peripheralist mold, certainly the model emerging today corrects this bias and places a new emphasis on the role of central factors in perception. Most important of all, I have the impression that the neurological model of perception that is now emerging begins for the first time to conform to our knowledge of more complex forms of perception, both in humans and at the infrahuman level.

## REFERENCES

1. ADRIAN, E. D. *The physical background of perception*. Oxford: Clarendon Press, 1947.
2. ADRIAN, E. D. The physiological basis of perception. In E. D. Adrian et al. (Eds.), *Brain mechanisms and consciousness*. Oxford: Blackwell, 1954.
3. ADRIAN, E. D., & MORUZZI, G. Cited in E. D. Adrian, *The physical background of perception*. Oxford: Clarendon Press, 1947.
4. BÉKÉSY, G. v. Description of some mechanical properties of the organ of Corti. *J. acoust. Soc. Amer.*, 1953, 25, 770-785.
5. BREMER, F. In E. D. Adrian et al. (Eds.), *Brain mechanisms and consciousness*. Oxford: Blackwell, 1954.
6. BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.
7. BRUNER, J. S., & POSTMAN, L. On the perception of incongruity: a paradigm. *J. Pers.*, 1949, 18, 206-223.
8. CRAIK, K. *The nature of explanation*. Cambridge, England: Cambridge University Press, 1952.
9. DAVIS, H. Psychophysiology of hearing and deafness. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
10. DETHIER, V. G. The dioptric apparatus of lateral ocelli. II. Visual capacities of the ocellus. *J. cell. comp. Physiol.*, 1943, 22, 115-126.
11. DITCHBURN, R. W. Report to the Experimental Psychology Group. Reading, England, January, 1957.
12. EXNER, S. *Die Physiologie der facettierten Augen von Krebsen und Insekten*. Leipzig und Wien, 1891.
13. FECHNER, G. T. *Elemente der Psychophysik*, I and II. Leipzig, 1869.
14. FESSARD, A. In E. D. Adrian et al. (Eds.), *Brain mechanisms and consciousness*. Oxford: Blackwell, 1954.
15. FESSARD, A. Mechanisms of nervous integration and conscious experience. In E. D. Adrian et al. (Eds.), *Brain mechanisms and consciousness*. Oxford: Blackwell, 1954.
16. FLETCHER, H. *Speech and hearing*. New York: D. Van Nostrand, 1929.
17. GALAMBOS, R. Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea. *J. Neurophysiol.*, 1956, 19, 424-431.
18. GALAMBOS, R., & DAVIS, H. The response of single auditory-nerve fibers to acoustic stimulation. *J. Neurophysiol.*, 1943, 6, 39-58.
19. GALAMBOS, R., & DAVIS, H. Action potentials from single auditory nerve fibers? *Science*, 1948, 108, 513.
20. GALAMBOS, R., SHEATZ, G., & VERNIER, V. G. Electrophysiological correlates of a conditioned response in cats. *Science*, 1956, 123, 376-377.
21. GIBSON, J. J. *The perception of the visual world*. Boston: Houghton Mifflin, 1950.

22. GRANIT, R. *Receptors and sensory perception*. New Haven: Yale Univ. Press, 1955.
23. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
24. HEBB, D. O. The role of neurological ideas in psychology. *J. Pers.* 1951, 20, 40.
25. HEBB, D. O. The problem of consciousness and introspection. In E. D. Adrian *et al.* (Eds.), *Brain mechanisms and consciousness*. Oxford: Blackwell, 1954.
26. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, 55, 297-313.
27. HERNANDEZ-PÉON, R., & SCHERRER, H. *Federation Proc.*, 1955, 14, 71.
28. HERNANDEZ-PÉON, R., SCHERRER, H., & JOUVET, M. Modification of electric activity in the cochlear nucleus during "attention" in unanesthetized cats. *Science*, 1956, 123, 331-332.
29. JAKOBSON, R., & HALLÉ, M. *Fundamentals of language*. Mouton: 'S-Gravenhagh (Holland), 1956.
30. KLÜVER, H. *Behavior mechanisms in monkeys*. Chicago: Univ. Chicago Press, 1933.
31. KUFFLER, S. W., & HUNT, C. C. The mammalian small nerve fibers: a system for efferent nervous regulation of muscle spindle discharge. *Proc. Assoc. Res. Nerv. Ment. Dis.*, 1952, Vol. 30.
32. LASHLEY, K. S. In search of the engram. *Symp. Soc. Exp. Biol.*, 1950, 4, 454-482.
33. LASHLEY, K. S. Dynamic processes in perception. In E. D. Adrian *et al.* (Eds.), *Brain mechanisms and consciousness*. Oxford: Blackwell, 1954.
34. LEKSELL, L. The action potential and excitatory effects of the small ventral root fibers to skeletal muscles. *Acta Physiol. Scand.*, 1945, 10, Suppl. 31.
35. MANDLER, G. Transfer of training as a function of degree of response overlearning. *J. exp. Psychol.*, 1954, 47, 411-417.
36. MCGILL, W. J., & ROSENBLITH, W. A. Electrical responses to two clicks: a simple statistical interpretation. *Bull. math. Biophysics*, 1951, 13, 69.
37. MILLER, G. A. Human memory and the storage of information. *Trans. on Information Theory, Inst. Radio Engrs.*, 1956, Vol. IT2, No. 3, 129-137.
38. PENFIELD, W. Studies of the cerebral cortex of man—a review and an interpretation. In E. D. Adrian *et al.* (Eds.), *Brain mechanisms and consciousness*. Oxford: Blackwell, 1954.
39. PFAFFMANN, C. Gustatory afferent impulses. *J. cell. comp. Physiol.*, 1941, 17, 243-258.
40. PIAGET, J. *Construction of reality in the child*. New York: Basic Books, 1954.
41. RIGGS, L. A. Visual acuity and the normal tremor of the eyes. *Science*, 1951, 114, 17-18.
42. ROSENBLITH, W. A. Electrical responses from the auditory nervous system. *Ann. Otolaryngology & Rhinology*, 1954, 63, 1-22.
43. SAUSSURE, F. DE. *Cours de linguistique general* (2nd ed.). Paris: Payot, 1922.
44. SENDEN, M. v. *Raum- und Gestaltauffassung bei operierten Blindgeborenen vor und nach der Operation*. Leipzig: Barth, 1932.
45. SHARPLESS, S., & JASPER, H. Habituation of the arousal reaction. *Brain*, 1956, 79, 655-680.
46. STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, 64, 153-181.
47. TASAKI, I. Nerve impulses in the individual auditory nerve fibers of guinea pig. *J. Neurophysiol.*, 1954, 17, 97-122.
48. THORPE, W. H. *Learning and instinct in animals*. Cambridge: Harvard Univ. Press, 1956.
49. THOULESS, R. H. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, 22, 216-241.
50. TUNTURI, A. R. Audio frequency localization in the acoustic cortex of the dog. *Amer. J. Physiol.*, 1944, 141, 397-403.
51. TUNTURI, A. R. Further afferent connections to the acoustic cortex of the dog. *Amer. J. Physiol.*, 1945, 144, 389-394.
52. ULLYOT, P. The behaviour of *Dendrocoelum lacteum*. II. Responses in non-directional gradients. *J. exp. Biol.*, 1936, 13, 265-278.
53. WALLACH, H. Brightness constancy and the nature of achromatic colors. *J. exp. Psychol.*, 1948, 38, 310-324.
54. WEBER, E. H. In E. G. Boring, *Sensation and perception in the history of experimental psychology*. New York: D. Appleton-Century, 1942. Pp. 495 ff.
55. WOODWORTH, R. S. Reinforcement of perception. *Amer. J. Psychol.*, 1947, 60, 119-124.

(Received February 18, 1957)

## MOTIVATIONAL DETERMINANTS OF RISK-TAKING BEHAVIOR

JOHN W. ATKINSON<sup>1</sup>

*University of Michigan*

There are two problems of behavior which any theory of motivation must come to grips with. They may finally reduce to one; but it will simplify the exposition which follows to maintain the distinction in this paper. The first problem is to account for an individual's selection of one path of action among a set of possible alternatives. The second problem is to account for the amplitude or vigor of the action tendency once it is initiated, and for its tendency to persist for a time in a given direction. This paper will deal with these questions in a conceptual framework suggested by research which has used thematic apperception to assess individual differences in strength of achievement motivation (1, 14, 15).

The problem of selection arises in experiments which allow the individual to choose a task among alternatives that differ in difficulty (level of aspiration). The problem of accounting for the vigor of response arises in studies which seek to relate individual differences in strength of motivation to the level of performance when response output at a particular task is the dependent variable. In treating these two problems, the discussion will be constantly focused on the relationship of achievement motivation to risk-taking behavior, an important association uncovered by McClelland (14) in the investigation of the role of achievement motivation in entre-

preneurship and economic development.<sup>2</sup>

Earlier studies have searched for a theoretical principle which would explain the relationship of strength of motive, as inferred from thematic apperception, to overt goal-directed performance. The effect of situation cues (e.g., of particular instructions) on this relationship was detected quite early (1), and subsequent experiments have suggested a theoretical formulation similar to that presented by Tolman (21) and Rotter (20). It has been proposed that *n* Achievement scores obtained from thematic apperception are indices of individual differences in the strength of achievement motive, conceived as a relatively stable disposition to strive for achievement or success. This motive-disposition is presumed to be latent until aroused by situation cues which indicate that some performance will be instrumental to achievement. The strength of *aroused* motivation to achieve as manifested in performance has been viewed as a function of both the strength of motive and the *expectancy* of goal-attainment aroused by situation cues. This conception has provided a fairly adequate explanation of experimental results to date, and several of its implications have been tested (1, 2).

The similarity of this conception to the expectancy principle of performance developed by Tolman, which also takes account of the effects of a third variable, *incentive*, suggested the need for experiments to isolate the effects on motiva-

<sup>1</sup> I wish to acknowledge the stimulation and criticism of colleagues at the Center for Advanced Study in the Behavioral Sciences (1955-56), and also the current support for this research by a grant from the Ford Foundation.

<sup>2</sup> McClelland, D. C. Interest in risky occupations among subjects with high achievement motivation. Unpublished paper, Harvard University, June, 1956.

tion of variations in strength of expectancy of success and variations in the incentive value of particular accomplishments. The discussion which follows was prompted by the results of several exploratory experiments. It represents an attempt to state explicitly how individual differences in the strength of achievement-related motives influence behavior in competitive achievement situations. A theoretical model will be presented first, then a brief summary of some as yet unpublished experimental evidence will be introduced in order to call the reader's attention to the kinds of research problems it raises and the scope of its implications.

Three variables require definition and, ultimately, independent measurement. The three variables are *motive*, *expectancy*, and *incentive*. Two of these—expectancy and incentive—are similar to variables presented by Tolman (21) and Rotter (20). An expectancy is a cognitive anticipation, usually aroused by cues in a situation, that performance of some act will be followed by a particular consequence. The strength of an expectancy can be represented as the subjective probability of the consequence, given the act.

The incentive variable has been relatively ignored, or at best crudely defined, in most research. It represents the relative attractiveness of a specific goal that is offered in a situation, or the relative unattractiveness of an event that might occur as a consequence of some act. Incentives may be manipulated experimentally as, for example, when amount of food (reward) or amount of shock (punishment) is varied in research with animals.

The third variable in this triumvirate—motive—is here conceived differently than, for example, in the common conception of motivation as nondirective but energizing *drive* (3). A motive is conceived as a disposition to strive for

a certain kind of satisfaction, as a capacity for satisfaction in the attainment of a certain class of incentives. The names given motives—such as achievement, affiliation, power—are really names of classes of incentives which produce essentially the same kind of experience of satisfaction: pride in accomplishment, or the sense of belonging and being warmly received by others, or the feeling of being in control and influential. McClelland (13, pp. 341–352 and 441–458; 15) has presented arguments to support the conception of motives as relatively general and stable characteristics of the personality which have their origins in early childhood experience. The idea that a motive may be considered a *capacity for satisfaction* is suggested by Winterbottom's (15, 22) finding that children who are strong in achievement motive are rated by teachers as deriving more pleasure from success than children who are weak in achievement motive.

The general aim of one class of motives, usually referred to as appetites or approach tendencies, is to maximize satisfaction of some kind. The achievement motive is considered a disposition to approach success.

The aim of another class of motives is to minimize pain. These have been called aversions, or avoidant tendencies. An avoidance motive represents the individual's capacity to experience pain in connection with certain kinds of negative consequences of acts. The motive to avoid failure is considered a disposition to avoid failure and/or a capacity for experiencing shame and humiliation as a consequence of failure.

*The principle of motivation.* The strength of motivation to perform some act is assumed to be a multiplicative function of the strength of the motive, the expectancy (subjective probability) that the act will have as a consequence the attainment of an incentive, and the



value of the incentive:  $\text{Motivation} = f(\text{Motive} \times \text{Expectancy} \times \text{Incentive})$ . This formulation corresponds to Tolman's (21) analysis of performance except, perhaps, in the conception of a motive as a relatively stable disposition. When both motivation to approach and motivation to avoid are simultaneously aroused, the resultant motivation is the algebraic summation of approach and avoidance. The act which is performed among a set of alternatives is the act for which the resultant motivation is most positive. The magnitude of response and the persistence of behavior are functions of the strength of motivation to perform the act relative to the strength of motivation to perform competing acts.

Recent experiments (2) have helped to clarify one problem concerning the relationship between measures of the strength of a particular motive ( $n$  Achievement) and performance. Performance is positively related to the strength of a particular motive only when an expectancy of satisfying that motive through performance has been aroused, and when expectancies of satisfying other motives through the same action have not been sufficiently aroused to confound the simple relationship. This is to say no more than that, when expectancies of attaining several different kinds of incentives are equally *salient* in a situation, the determination of motivation to perform an act is very complex. Performance is then overdetermined in the sense that its strength is now a function of the several different kinds of motivation which have been aroused. The *ideal situation* for showing the relationship between the strength of a particular motive and behavior is one in which the only *reason* for acting is to satisfy that motive.

The theoretical formulation which follows pertains to such an *ideal achievement-related situation*, which is at best

only approximated in actual experimentation or in the normal course of everyday life. The discussion will deal only with the effects of the two motives, to achieve and to avoid failure, normally aroused whenever performance is likely to be evaluated against some standard of excellence.

*Behavior directed toward achievement and away from failure.* The problem of selection is confronted in the level-of-aspiration situation where the individual must choose among tasks which differ in degree of difficulty. The problem of accounting for the vigor of performance arises in the situation which will be referred to as *constrained performance*. Here there is no opportunity for the individual to choose his own task. He is simply given a task to perform. He must, of course, decide to perform the task rather than to leave the situation. There is a problem of selection. In referring to this situation as constrained performance, it is the writer's intention to deal only with those instances of behavior in which motivation for the alternative of leaving the situation is less positive or more negative than for performance of the task that is presented. Hence, the individual does perform the task that is given. The level of performance is the question of interest.

Elaboration of the implications of the multiplicative combination of motive, expectancy, and incentive, as proposed to account for strength of motivation, will be instructive if we can find some reasonable basis for assigning numbers to the different variables. The strength of expectancy can be represented as a subjective probability ranging from 0 to 1.00. But the problem of defining the positive incentive value of a particular accomplishment and the negative incentive value of a particular failure is a real stickler.

In past discussions of level of aspiration, Escalona and Festinger (see 12)

have assumed that, within limits, the attractiveness of success is a positive function of the difficulty of the task, and that the unattractiveness of failure is a negative function of difficulty, when the type of activity is held constant. The author will go a few steps farther with these ideas, and assume that degree of difficulty can be inferred from the subjective probability of success ( $P_s$ ). The task an individual finds difficult is one for which his subjective probability of success ( $P_s$ ) is very low. The task an individual finds easy is one for which his subjective probability of success ( $P_s$ ) is very high. Now we are in a position to make simple assumptions about the incentive values of success or failure at a particular task. Let us assume that the incentive value of success ( $I_s$ ) is a positive linear function of difficulty. If so, the value  $1 - P_s$  can represent  $I_s$ , the incentive value of success. When  $P_s$  is high (e.g., .90), an easy task,  $I_s$  is low (e.g., .10). When  $P_s$  is low (e.g., .10), a difficult task,  $I_s$  is high (e.g., .90). The negative incentive value of failure ( $I_f$ ) can be taken as  $-P_s$ . When  $P_s$  is high (e.g., .90), as in confronting a very easy task, the sense of humiliation accompanying failure is also very great (e.g.,  $-.90$ ).

However, when  $P_s$  is low (e.g., .10), as in confronting a very difficult task, there is little embarrassment in failing (e.g.,  $-.10$ ). We assume, in other words, that the (negative) incentive value of failure ( $I_f$ ) is a negative linear function of difficulty. It is of some importance to recognize the dependence of incentive values intrinsic to achievement and failure upon the subjective probability of success. One cannot anticipate the thrill of a great accomplishment if, as a matter of fact, one faces what seems a very easy task. Nor does an individual experience only a minor sense of pride after some extraordinary feat against what seemed to him overwhelming odds. The implications of the scheme which follows rest heavily upon the assumption of such a dependence.

In Table 1, values of 1 have been arbitrarily assigned to the achievement motive ( $M_s$ ) and the motive to avoid failure ( $M_f$ ). Table 1 contains the strength of motivation to approach success ( $M_s \times P_s \times I_s$ ) and motivation to avoid failure ( $M_f \times P_f \times I_f$ ) through performance of nine different tasks labeled A through I. The tasks differ in degree of difficulty as inferred from the subjective probability of success ( $P_s$ ). The incentive values of success and fail-

TABLE 1  
AROUSAL MOTIVATION TO ACHIEVE (APPROACH) AND TO AVOID FAILURE (AVOIDANCE) AS A  
JOINT FUNCTION OF MOTIVE ( $M$ ), EXPECTANCY ( $P$ ), AND INCENTIVE ( $I$ ),  
WHERE  $I_s = (1 - P_s)$  AND  $I_f = (-P_s)$

	Motivation to Achieve				Motivation to Avoid Failure				Resultant Motivation (Approach—Avoidance)
	$M_s \times P_s \times I_s = \text{Approach}$				$M_f \times P_f \times I_f = \text{Avoidance}$				
Task A	1	.10	.90	.09	1	.90	-.10	-.09	0
Task B	1	.20	.80	.16	1	.80	-.20	-.16	0
Task C	1	.30	.70	.21	1	.70	-.30	-.21	0
Task D	1	.40	.60	.24	1	.60	-.40	-.24	0
Task E	1	.50	.50	.25	1	.50	-.50	-.25	0
Task F	1	.60	.40	.24	1	.40	-.60	-.24	0
Task G	1	.70	.30	.21	1	.30	-.70	-.21	0
Task H	1	.80	.20	.16	1	.20	-.80	-.16	0
Task I	1	.90	.10	.09	1	.10	-.90	-.09	0

ure at each of the tasks have been calculated directly from the assumptions that incentive value of success equals  $1 - P_s$  and that incentive value of failure equals  $-P_s$ ; and  $P_s$  and  $P_f$  are assumed to add to 1.00.

Table 1 may be considered an extension of ideas presented in the *resultant valence* theory of level of aspiration by Escalona and Festinger (12). The present formulation goes beyond their earlier proposals (a) in making specific assumptions regarding the incentive values of success and failure, and (b) in stating explicitly how individual differences in strength of achievement motive and motive to avoid failure influence motivation.<sup>3</sup>

When the achievement motive is stronger ( $M_s > M_f$ ). The right-hand column of Table 1 shows the resultant motivation for each of the tasks in this special case where achievement motive and motive to avoid failure are equal in strength. In every case there is an approach-avoidance conflict with resultant motivation equal to 0. This means that if the achievement motive were stronger than the motive to avoid failure—for example, if we assigned  $M_s$  a value of 2—the resultant motivation would become positive for each of the tasks and its magnitude would be the same as in the column labeled *Approach*. Let us therefore consider only the strength of

approach motivation for each of the tasks, to see the implications of the model for the person in whom the need for achievement is stronger than his disposition to avoid failure.

One thing is immediately apparent. Motivation to achieve is strongest when uncertainty regarding the outcome is greatest, i.e., when  $P_s$  equals .50. If the individual were confronted with all of these tasks and were free to set his own goal, he should choose Task E where  $P_s$  is .50, for this is the point of maximum approach motivation. The strength of motivation to approach decreases as  $P_s$  increases from .50 to near certainty of success ( $P_s = .90$ ), and it also decreases as  $P_s$  decreases from .50 to near certainty of failure ( $P_s = .10$ ).

If this person were to be confronted with a single task in what is here called the constrained performance situation, we should expect him to manifest strongest motivation in the performance of a task of intermediate difficulty where  $P_s$  equals .50. If presented either more difficult tasks or easier tasks, the strength of motivation manifested in performance should be lower. The relationship between strength of motivation as expressed in performance level and expectancy of success at the task, in other words, should be described by a bell-shaped curve.

When the motive to avoid failure is stronger ( $M_f > M_s$ ). Let us now ignore the strength of approach motivation and tentatively assign it a value of 0, in order to examine the implications of the model for any case in which the motive to avoid failure is the stronger motive. The resultant motivation for each task would then correspond to the values listed in the column labeled *Avoidance*.

What should we expect of the person in whom the disposition to avoid failure is stronger than the motive to achieve? It is apparent at once that

<sup>3</sup> In the resultant valence theory of level of aspiration, the resultant force ( $f^*$ ) for a particular level of difficulty equals probability of success ( $P_s$ ) times valence of success ( $Va_s$ ) minus probability of failure ( $P_f$ ) times valence of failure ( $Va_f$ ). It is assumed that the valence of a goal [ $Va(G)$ ] depends partly on the properties of the activity and specific goal ( $G$ ) and partly on the state of need [ $t(G)$ ] of the person, [ $Va(G) = F(G, t(G))$ ] (11, p. 273). In the present conception, the relative rewarding or punishing properties of specific goals (i.e., incentives) and the more general disposition of the person toward a class of incentives (i.e., his motive) are given independent status.

the resultant motivation for every task would be negative for him. This person should want to avoid all of the tasks. Competitive achievement situations are unattractive to him. If, however, he is constrained (e.g., by social pressures) and asked to set his level of aspiration, he should *avoid* tasks of intermediate difficulty ( $P_s = .50$ ) where the arousal of anxiety about failure is greatest. He should choose either the easiest ( $P_s = .90$ ) or the most difficult task ( $P_s = .10$ ). The strength of avoidant motivation is weakest at these two points.

In summary, the person in whom the achievement motive is stronger should set his level of aspiration in the intermediate zone where there is moderate risk. To the extent that he has any motive to avoid failure, this means that he will voluntarily choose activities that *maximize* his own anxiety about failure! On the other hand, the person in whom the motive to avoid failure is stronger should select either the easiest of the alternatives or should be extremely speculative and set his goal where there is virtually no chance for success. These are activities which *minimize* his anxiety about failure.

How does the more fearful person behave when offered only a specific task to perform? He can either perform the task or leave the field. If he chooses to leave the field, there is no problem. But if he is constrained, as he must be to remain in any competitive achievement situation, he will stay at the task and presumably work at it. But how hard will he work at it? He is motivated to avoid failure, and when constrained, there is only one path open to him to avoid failure—success at the task he is presented. So we expect him to manifest the strength of his motivation to avoid failure in performance of the task. He, too, in other words, should

*try hardest*<sup>4</sup> when  $P_s$  is .50 and less hard when the chance of winning is either greater or less. The 50–50 alternative is the last he would choose if allowed to set his own goal, but once constrained he must try hard to avoid the failure which threatens him. Not working at all will guarantee failure of the task. Hence, the thought of not working at all should produce even stronger avoidant motivation than that aroused by the task itself.

In other words, irrespective of whether the stronger motive is to achieve or to avoid failure, the strength of motivation to perform a task when no alternatives are offered and when the individual is constrained should be greatest when  $P_s$  is .50. This is the condition of greatest uncertainty regarding the outcome. But when there are alternatives which differ in difficulty, the choice of level of aspiration by persons more disposed to avoid failure is diametrically opposite to that of persons more disposed to seek success. The person more motivated to achieve should prefer a moderate risk. His level of aspiration will fall at the point where his positive motivation is strongest, at the point where the odds seem to be 50–50. The fearful person, on the other hand, must select a task even though all the alternatives are threatening to him. He prefers the least threatening of the available alternatives: either the task which is so easy he cannot fail, or the task which is so difficult that failure would be no cause for self-blame and embarrassment.

The tendency for anxious persons to set either extremely high or very low aspirations has been noted over and over

<sup>4</sup>I do not mean to exclude the possibility that the very anxious person may suffer a performance decrement due to the arousal of some "task-irrelevant" avoidant responses, as proposed in the interpretation of research which has employed the Mandler-Sarason Measure of Test Anxiety (16).

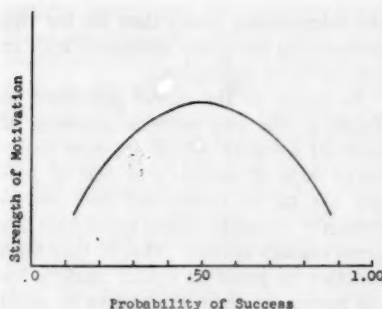


FIG. 1. Strength of motivation to achieve or to avoid failure as a function of the subjective probability of success, i.e., the difficulty of the task.

again in the literature on level of aspiration (12). Typically, groups of persons for whom the inference of greater anxiety about failure seems justified on the basis of some personality assessment show a much greater variance in level of aspiration than persons whose motivation is inferred to be more normal or less anxious. When the details of behavior are examined, it turns out that they are setting their aspiration level either *defensively* high or *defensively* low.

Without further assumptions, the theory of motivation which has been presented when applied to competitive-achievement activity implies that the relationship of constrained performance to expectancy of goal-attainment should take the bell-shaped form shown in Fig. 1, whether the predominant motive is to achieve or to avoid failure. Further, the theory leads to the prediction of exactly opposite patterns for setting the level of aspiration when the predominant motivation is approach and when it is avoidant, as shown in Fig. 2.

Both of these hypotheses have been supported in recent experiments. The writer<sup>5</sup> offered female college students

<sup>5</sup> Atkinson, J. W. Towards experimental analysis of human motivation in terms of

a modest monetary prize for good performance at two 20-minute tasks. The probability of success was varied by instructions which informed the subject of the number of persons with whom she was in competition and the number of monetary prizes to be given. The stated probabilities were  $\frac{1}{20}$ ,  $\frac{1}{3}$ ,  $\frac{1}{2}$ , and  $\frac{3}{4}$ . The level of performance was higher at the intermediate probabilities than at the extremes for subjects having high thematic apperceptive *n* Achievement scores, and also for subjects who had low *n* Achievement scores, presumably a more fearful group.

McClelland<sup>6</sup> has shown the diametrically opposite tendencies in choice of level of aspiration in studies of children in kindergarten and in the third grade. One of the original level-of-aspiration experiments, the ring-toss experiment, was repeated with five-year-olds, and a non-verbal index of the strength of achieve-

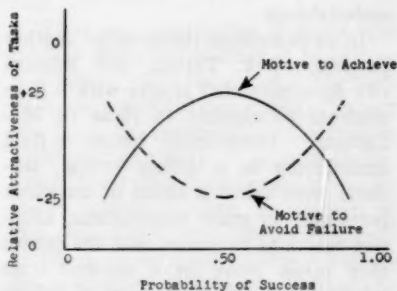


FIG. 2. Relative attractiveness of tasks which differ in subjective probability of success (i.e., in difficulty). The avoidance curve has been inverted to show that very difficult and very easy tasks arouse less fear of failure and hence are less unattractive than moderately difficult tasks.

motives, expectancies, and incentives. To appear in *Motives in fantasy, action, and society*. Princeton: Van Nostrand (in preparation).

<sup>6</sup> McClelland, D. C. Risk-taking in children with high and low need for achievement. To appear in *Motives in fantasy, action, and society*. Princeton: Van Nostrand (in preparation).



ment motive was employed. Children who were high in  $n$  Achievement more frequently set their level of aspiration in the intermediate range of difficulty. They took more shots from a modest distance. Children who were low in  $n$  Achievement showed a greater preponderance of choices at the extreme levels of difficulty. They more often stood right on top of the peg or stood so far away that success was virtually impossible. The same difference between high and low  $n$  Achievement groups was observed on another task with children in the third grade. McClelland views these results as consistent with his theoretical argument concerning the role of achievement motivation in entrepreneurship and economic development (14). He has called attention to the relationship between achievement motivation and an interest in enterprise which requires moderate or calculated risks, rather than very safe or highly speculative undertakings.

In an experiment designed for another purpose, Clark, Teevan, and Ricciuti (4) have presented results with college students comparable to those of McClelland. Immediately before a final examination in a college course, students were asked a series of questions pertaining to grade expectations, affective reactions to grades, and the grades they would *settle for* if excused from taking the exam. A number of indices were derived from responses to these questions, by which the students were classified as: *hopeful of success*, i.e., if the *settle-for* grade was near the maximum grade the student thought he could possibly achieve; *fearful of failure*, i.e., if the *settle-for* grade was near the minimum grade the student thought he might possibly drop to; and *intermediate*, i.e., if the *settle-for* grade fell somewhere between these two extremes. Previously obtained  $n$  Achievement scores were significantly higher for

the *intermediate* group than for the two groups who set either extremely high or low levels of aspiration.

In terms of the model presented in Table 1, the two extreme patterns of aspirant behavior which are here designated *hope of success* and *fear of failure* are to be considered two *phenotypically* dissimilar alternatives that are *genotypically* similar. That is, they both function to avoid or reduce anxiety for the person in whom the motive to avoid failure is stronger than the motive to achieve.

A question may arise concerning the legitimacy of inferring relatively stronger motive to avoid failure from a low  $n$  Achievement score in thematic apperception. The inference seems justified on several counts. First, the kind of learning experience which is thought to contribute to the development of a positive motive to achieve (15, 22) seems incompatible with the kind of experience which would contribute to the development of an avoidant motive. In any specific early learning experience in which successful independent accomplishment is encouraged and rewarded, it seems impossible for incompetence, at the same time, to be punished. Second, even if it is assumed that high and low  $n$  Achievement groups may be equal in the disposition to be fearful of failure, the fact that one group does not show evidence of a strong motive to achieve (the group with low  $n$  Achievement scores) suggests that fear of failure should be *relatively* stronger in that group than in the group which does show evidence of strong  $n$  Achievement (high  $n$  Achievement scores). Finally, Raphelson (19) has presented evidence that  $n$  Achievement, as measured in thematic apperception, is *negatively* related to both scores on the Mandler-Sarason Scale of Test Anxiety and a psychogalvanic index of manifest anxiety obtained in a test situation. Test

anxiety scores and the psychogalvanic index of manifest anxiety were *positively* correlated, as they should be if each is an effective measure of fear aroused in a competitive situation.

Although a low *n* Achievement score can hardly be viewed as a direct index of the disposition to avoid failure, there seems good presumptive evidence that fear of failure is *relatively* stronger than the achievement motive in such a group. And this presumption is all the theory demands to explain the pattern of goal setting which focuses upon the extremes in the range of difficulty among persons low in *n* Achievement.

The details of the exploratory experiments suggest that one further assumption be made. In both experiments, the high *n* Achievement groups showed evidence of maximum motivation when the observed or stated probability of success was approximately .33. At this point, the high *n* Achievement group showed the highest level of constrained performance. And this point was most favored by the high *n* Achievement group in setting level of aspiration in the McClelland experiment. The assumption to be made seems a reasonable one: the relative strength of a motive influences the subjective probability of the consequence consistent with that motive—i.e., biases it upwards. In other words, the stronger the achievement motive relative to the motive to avoid failure, the higher the subjective probability of success, given stated odds. The stronger the motive to avoid failure relative to the achievement motive, the higher the subjective probability of failure, given stated odds or any other objective basis for inferring the strength of expectancy. Some evidence from two earlier studies is pertinent. When subjects stated the score that they *expected* to make on a test with very ambiguous or conflicting cues from past performance (15, p. 247) or when faced with a novel task at

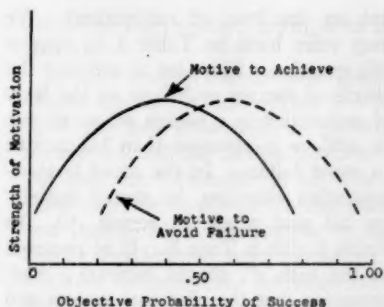


FIG. 3. Strength of motivation to achieve and to avoid failure as a function of the *objective* probability of success. It is assumed that the subjective probability of the consequence consistent with the stronger motive is biased upwards.

which they had no experience (18), the stated level of *expectation* was positively related to *n* Achievement. The biasing effect of the motive on subjective probability should diminish with repeated learning experience in the specific situation.

When this assumption is made, the point of maximum motivation to achieve now occurs where the stated (objective) odds are somewhat *lower* than .50; and the point of maximum motivation to avoid failure occurs at a point somewhat higher than stated odds of .50, as shown in Fig. 3. The implications of this assumption for constrained performance in somewhat novel situations are evident in the figure. When the achievement motive is stronger than the motive to avoid failure, there should be a tendency for stronger motivation to be expressed in performance when the objective odds are long, i.e., below .50. When the motive to avoid failure is stronger than the achievement motive, there should be greater motivation expressed when the objective odds are short, i.e., above .50.

*The effects of success and failure.* Let us return to the model and ask, What are the effects of success and fail-

ure on the level of motivation? We may refer back to Table 1 to answer this question. First, let us consider the effects of success or failure on the level of motivation in a person whose motive to achieve is stronger than his motive to avoid failure. In the usual level-of-aspiration situation, he should initially set his goal where  $P_s$  equals .50. In Table 1, this is Task E. If he succeeds at the task,  $P_s$  should increase. And, assuming that the effects of success and failure generalize to similar tasks, the  $P_s$  at Task D which was initially .40 should increase toward .50. On the next trial,  $P_s$  at Task E is now greater than .50, and  $P_s$  at Task D now approaches .50. The result of this change in  $P_s$  is diminished motivation to achieve at the old task, E, and increased motivation to achieve at Task D, an *objectively more difficult task*. The observed level of aspiration should increase in a step-like manner following success, because there has been a change in motivation.

A further implication of the change in strength of motivation produced by the experience of success is of great consequence: given a single, very difficult task (e.g.,  $P_s = .10$ ), the effect of continued success in repeated trials is first a gradual increase in motivation as  $P_s$  increases to .50, followed by a gradual decrease in motivation as  $P_s$  increases further to the point of certainty ( $P_s = 1.00$ ). Ultimately, as  $P_s$  approaches 1.00, satiation or loss of interest should occur. The task no longer arouses any motivation at all. Why? Because the subjective probability of success is so high that the incentive value is virtually zero. Here is the clue to understanding how the achievement motive can remain insatiable while satiation can occur for a particular line of activity. The strength of motive can remain unchanged, but interest in a particular task can diminish com-

pletely. Hence, when free to choose, the person who is stronger in achievement motive should always look for new and more difficult tasks as he masters old problems. If constrained, the person with a strong achievement motive should experience a gradual loss of interest in his work. If the task is of intermediate difficulty to start with ( $P_s = .50$ ), or is definitely easy ( $P_s > .50$ ), his interest should begin to wane after the initial experience of success.

But what of the effect of failure on the person who is more highly motivated to achieve than to avoid failure? Once more we look at the *Approach* column of Table 1. If he has chosen Task E ( $P_s = .50$ ) to start with and fails at it, the  $P_s$  is reduced. Continued failure will mean that soon Task F (formerly  $P_s = .60$ ) will have a  $P_s$  near .50. He should shift his interest to this task, which was *objectively less difficult* in the initial ordering of tasks. This constitutes what has been called a lowering of the level of aspiration. He has moved to the easier task as a consequence of failure.

What is the effect of continued failure at a single task? If the initial task is one that appeared relatively easy to the subject (e.g.,  $P_s = .80$ ) and he fails, his motivation should increase! The  $P_s$  will drop toward .70, but the incentive value or attractiveness of the task will increase. Another failure should increase his motivation even more. This will continue until the  $P_s$  has dropped to .50. Further failure should then lead to a gradual weakening of motivation as  $P_s$  decreases further. In other words, the tendency of persons who are relatively strong in achievement motive to persist at a task in the face of failure is probably attributable to the relatively high subjective probability of success, initially. Hence, failure has the effect of increasing the strength of their motivation, at least for a time. Ultimately,

however, interest in the task will diminish if there is continued failure. If the initial task is perceived by the person as very difficult to start with ( $P_s < .50$ ), motivation should begin to diminish with the first failure.

Let us turn now to the effect of success and failure on the motivation of the person who is more strongly disposed to be fearful of failure. If the person in whom the motive to avoid failure is stronger has chosen a very difficult task in setting his level of aspiration (e.g., Task A where  $P_s = .10$ ) and succeeds,  $P_s$  increases and his motivation to avoid the task is paradoxically increased! It would almost make sense for him deliberately to fail, in order to keep from being faced with a stronger threat on the second trial. If there are more difficult alternatives, he should raise his level of aspiration to avoid anxiety! Fortunately for this person, his strategy (determined by the nature of his motivation) in choosing a very difficult task to start with protects him from this possibility, because  $P_s$  is so small that he will seldom face the paradoxical problem just described. If he fails at the most difficult task, as is likely,  $P_s$  decreases further,  $P_f$  increases further, and the aroused motivation to avoid failure is reduced. By continued failure he further reduces the amount of anxiety about failure that is aroused by this most difficult task. Hence, he should continue to set his level at this point. If he plays the game long enough and fails continuously, the probability of failure increases for all levels of difficulty. Sooner or later the minimal motivation to avoid failure at the most difficult task may be indistinguishable from the motivation to avoid failure at the next most difficult task. This may ultimately allow him to change his level of aspiration to a somewhat less difficult task without acting in gross contra-

diction to the proposed principle of motivation.

If our fearful subject has initially chosen the easiest task (Task I where  $P_s = .90$ ) and if he fails,  $P_s$  decreases toward .80, and his motivation to avoid the task also increases. If there is no easier task, the most difficult task should now appear least *unattractive* to him, and he should jump from the easiest to the most difficult task. In other words, continued failure at a very easy task decreases  $P_s$  toward .50; and, as Table 1 shows, a change of this sort is accompanied by increased arousal of avoidant motivation. A wild and apparently irrational jump in level of aspiration from very easy to very difficult tasks, as a consequence of failure, might be mistakenly interpreted as a possible effort on the part of the subject to gain social approval by seeming to set high goals. The present model predicts this kind of activity without appealing to some extrinsic motive. It is part of the strategy of minimizing expected pain of failure after one has failed at the easiest task.

If our fear-disposed subject is successful at the most simple task, his  $P_s$  increases, his  $P_f$  decreases, and his motivation to avoid this task decreases. The task becomes less and less unpleasant. He should continue playing the game with less anxiety.

Table 1, when taken in its entirety, deals with the special case of the person in whom the two motives are exactly equal in strength. The implications are clear. In the constrained-performance situation, he should work hardest when the probability of success is .50, because motivation to achieve and motivation to avoid failure will summate in the constrained instrumental act which is at the same time the pathway toward success and away from failure. (This summation should also occur in the cases where one motive is stronger.) But in the level-of-aspiration setting

where there is an opportunity for choice among alternatives, the avoidance motivation exactly cancels out the approach motivation. Hence, the resultant motivation for each of the alternatives is zero. His choice of level of aspiration cannot be predicted from variables intrinsic to the achievement-related nature of the task. If there is any orderly pattern in this conflicted person's level of aspiration, the explanation of it must be sought in extrinsic factors, e.g. *the desire to gain social approval*. Such a desire can also be conceptualized in terms of motive, expectancy, and incentive, and the total motivation for a particular task can then be attributed to both achievement-related motives and other kinds of motives engaged by the particular features of the situation.

In recent years there has been something of a rebirth of interest in the problems of level of aspiration, particularly in pathological groups. The tendency for anxious groups to show much greater variability in level of aspiration, setting their goals either very high or very low relative to less anxious persons, was noted in early studies by Sears, Rotter, and others (12). Miller (17), Himmelweit (9), and Eysenck and Himmelweit (8) have produced substantial evidence that persons with affective disorders (neurasthenia or dysthymia) typically set extremely high goals for themselves; hysterics, on the other hand, show a minimal level of aspiration, often setting their future goal even below the level of past performance. In all of these studies, normal control groups have fallen between these two extremes, as might be expected from the present model if *normals* are relatively more positive in their motivation in achievement-related situations.

In the work of Eysenck (7) and his colleagues, both dysthymics and hysterics show greater *neuroticism* than nor-

mal subjects. Eysenck's interpretation of this factor as autonomic sensitivity is consistent with the implications of the present model, which attributes the setting of extremely high or low levels of aspiration to relatively strong motivation to avoid failure. A second factor, *extraversion-introversion*, discriminates the affective disorders and hysterics where the present model, dealing only with motives intrinsic to the competitive achievement situation, does not. An appeal to some other motivational difference, e.g., in strength of *n* Affiliation, might also predict the difference in pattern of level of aspiration.

#### *Probability Preferences*

The present analysis is relevant to another domain of current research interest, that specifically concerned with the measurement of subjective probability and utility. Edwards (5, 6), for example, has reported probability preferences among subjects offered alternative bets having the same expected value. We<sup>†</sup> have repeated the Edwards type experiment (e.g., 6/6 of winning 30¢ versus 1/6 of winning \$1.80) with subjects having high and low *n* Achievement scores. The results show that persons high in *n* Achievement more often prefer intermediate probabilities (4/6, 3/6, 2/6) to extreme probabilities (6/6, 5/6, 1/6) than do persons low in *n* Achievement. What is more, the same differential preference for intermediate risk was shown by these *same* subjects when they were allowed to choose the distance from the target for their shots in a shuffleboard game. In other words, the incentive values of winning *qua* winning, and losing *qua* losing, presumably developed in achievement activities early in life, generalize to the gambling

<sup>†</sup> Atkinson, J. W., Bastian, J. R., Earl, R. W., and Litwin, G. H. The achievement motive, goal-setting, and probability preferences (in preparation).



situation in which winning is really *not* contingent upon one's own skill and competence.

### *Social Mobility Aspirations*

Finally, the present model may illuminate a number of interesting research possibilities having to do with social and occupational mobility. The ranking of occupations according to their prestige in Western societies clearly suggests that occupations accorded greater prestige are also more difficult to attain. A serious effort to measure the perceived probability of being able to attain certain levels on the occupational ladder should produce a high negative correlation with the usual ranking on prestige. If so, then the present model for level of aspiration, as well as its implications for persons who differ in achievement-related motives, can be applied to many of the sociological problems of mobility aspirations. A recent paper by Hyman (10) has laid the groundwork for such an analysis.

### SUMMARY

A theoretical model is presented to explain how the motive to achieve and the motive to avoid failure influence behavior in any situation where performance is evaluated against some standard of excellence. A conception of motivation in which strength of motivation is a joint multiplicative function of motive, expectancy (subjective probability), and incentive is offered to account for the selection of one task among alternatives which differ in difficulty (level of aspiration), and also to account for performance level when only one task is presented. It is assumed that the incentive value of success is a positive linear function of difficulty as inferred from the subjective probability of success; and negative incentive value of failure is assumed to be a negative linear function of difficulty. The major im-

plications of the theory are (a) that performance level should be greatest when there is greatest uncertainty about the outcome, i.e., when subjective probability of success is .50, whether the motive to achieve or the motive to avoid failure is stronger within an individual; but (b) that persons in whom the achievement motive is stronger should prefer intermediate risk, while persons in whom the motive to avoid failure is stronger should avoid intermediate risk, preferring instead either very easy and safe undertakings or extremely difficult and speculative undertakings. Results of several experiments are cited, and the implications of the theoretical model for research on probability preferences in gambling and studies of social mobility aspirations are briefly discussed.

### REFERENCES

1. ATKINSON, J. W. Explorations using imaginative thought to assess the strength of human motives. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation*, 1954. Lincoln, Nebr.: Univ. of Nebraska Press, 1954.
2. ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal-attainment. *J. abnorm. soc. Psychol.*, 1956, 53, 361-366.
3. BROWN, J. S. Problems presented by the concept of acquired drives. In *Current theory and research in motivation*. Lincoln, Nebr.: Univ. of Nebraska Press, 1953.
4. CLARK, R. A., TEEVAN, R., & RICCIUTI, H. N. Hope of success and fear of failure as aspects of need for achievement. *J. abnorm. soc. Psychol.*, 1956, 53, 182-186.
5. EDWARDS, W. Probability preferences in gambling. *Amer. J. Psychol.*, 1953, 66, 349-364.
6. EDWARDS, W. The theory of decision making. *Psychol. Bull.*, 1954, 51, 380-417.
7. EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, 101, 28-51.
8. EYSENCK, H. J., & HIMMELWEIT, H. T. An experimental study of the reactions

- of neurotics to experiences of success and failure. *J. gen. Psychol.*, 1946, 35, 59-75.
9. HIMMELWEIT, H. T. A comparative study of the level of aspiration of normal and neurotic persons. *Brit. J. Psychol.*, 1947, 37, 41-59.
  10. HYMAN, H. H. The value systems of different classes: a social psychological contribution to the analysis of stratification. In R. Bendix & S. M. Lipset (Eds.), *Class, status, and power*. Glencoe, Ill.: Free Press, 1953.
  11. LEWIN, K. *Field theory in social science*. D. Cartwright (Ed.). New York: Harper Bros., 1951.
  12. LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*, Vol. 1, Chap. 10. New York: Ronald Press, 1944.
  13. MCCLELLAND, D. C. *Personality*. New York: William Sloane Associates, 1951.
  14. MCCLELLAND, D. C. Some social consequences of achievement motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1955. Lincoln: Univer. of Nebraska Press, 1955.
  15. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
  16. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 16, 115-118.
  17. MILLER, D. R. Responses of psychiatric patients to threat of failure. *J. abnorm. soc. Psychol.*, 1951, 46, 378-387.
  18. POTTARST, B. C. The achievement motive and level of aspiration after experimentally induced success and failure. Unpublished doctor's dissertation, Univer. of Michigan, 1956.
  19. RAPHELSON, A. Imaginative and direct verbal measures of anxiety related to physiological reactions in the competitive achievement situation. Unpublished doctor's dissertation, Univer. of Michigan, 1956.
  20. ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
  21. TOLMAN, E. C. Principles of performance. *Psychol. Rev.*, 1955, 62, 315-326.
  22. WINTERBOTTOM, M. R. The relation of childhood training in independence to achievement motivation. Unpublished doctor's dissertation, Univer. of Michigan, 1952.

(Received February 7, 1957)

## DURABLE SECONDARY REINFORCEMENT: METHOD AND THEORY<sup>1</sup>

DONALD W. ZIMMERMAN

University of Illinois

For at least a decade now, the importance of secondary reinforcement has been increasingly emphasized by learning theorists and by systematists in general. It is true that Spence (22) has recently revised his earlier position (21) in which this phenomenon played an important role. And McClelland (11) has for some time questioned its importance because of its apparent instability. Nevertheless, secondary reinforcement is alluded to more and more frequently in the psychological literature; and an empirical demonstration of its reality and durability, if reality and durability it has, becomes correspondingly more desirable.

The purpose of the present paper is to report a method whereby secondary reinforcement can be made extraordinarily effective and stable. Some of the theoretical implications of this finding will also be briefly discussed.

### I. A TWO-STAGE INTERMITTENT RE- INFORCEMENT PROCEDURE WHICH LEADS TO POWERFUL SECONDARY REINFORCEMENT

With the experimental procedure to be reported here, it has been found possible to give a secondary reinforcer such strength that it could, if desired, be used as a class demonstration of the Law of Effect, much in the manner of such primary reinforcers as food and water. The method whereby such stable sec-

ondary reinforcement can be established is as follows:

*Training.* A thirsty rat is put into a small (11 × 11 × 11-in.) box fitted with a motor-driven water delivery mechanism. After the animal becomes habituated to the situation, a formerly neutral stimulus is given secondary reinforcement potential in this way: A two-second buzzer, followed by operation of the water delivery zipper, is presented at approximately one-minute intervals. Buzzer presentations are programmed independently of the behavior of the animal, with the exception that the buzzer never sounds while the animal hovers over the water-delivery aperture.

After the approach response has been firmly established, water reinforcements are omitted following the buzzer, at first on alternate presentations, then successively in longer runs. The ratio of rewarded to nonrewarded presentations varies somewhat randomly, the mean number of reinforcement omissions, however, moving upward. A 1:10 ratio is finally stabilized, with the longest single nonreinforced run being 1:14. What is involved here, then, is a *variable ratio* of water-rewarded to nonrewarded S<sup>r</sup> (secondary-reinforcement stimulus) presentations,<sup>2</sup> this ratio being gradually increased. During this training the vigor with which the animal leaps to the water delivery aperture following the buzzer does not lessen, in spite of the fact that the frequency of "pay off" is progressively decreasing.

<sup>1</sup> The author wishes to express his gratitude to Dr. O. H. Mowrer for offering many helpful suggestions during the course of the research here reported and for reading and criticizing the manuscript.

<sup>2</sup> For convenience the following symbols will be used: S<sup>r</sup> for secondary reinforcer, S<sup>D</sup> for positive discriminative stimulus, S<sup>A</sup> for negative discriminative stimulus.

During this type of training what happens is that the buzzer becomes a signal "that water *may* now be available." Water is received at no time *except when the buzzer sounds*, so the animal soon learns to "pay attention" to the buzzer—as it were, to "jump at the chance" which it affords. But the buzzer is not a sure sign of water.

**Testing.** Following the training procedure just described, the buzzer is tested for its secondary reinforcing properties by presenting it as a consequence of the rat's pressing a small bar which is now available. The buzzer ( $S^r$ ) alone is thus used to strengthen this response, no further water reinforcements being given. However, instead of having the buzzer invariably follow bar pressing, the procedure again involves intermittent reinforcement. Here various options regarding the detailed procedure that may be used are open to the experimenter. The one that produced the results shown in Fig. 1 went as follows. On the first day of testing, the bar-pressing behavior was established by reinforcing (with the buzzer) the first six responses (segment A). Thereafter  $S^r$  was given following a response only when at least a minute had elapsed since the preceding reinforcement (segment B). When the buzzer was discontinued entirely, a typical extinction performance followed (segment C). Reconditioning, with secondary reinforcement only, was possible on a second day, without intervening primary reinforcement of the secondary reinforcer. Various fixed-ratio schedules have also given satisfactory results.

A point to be stressed here is that the intermittent reinforcement schedules of the kind just suggested for use during the test period will increase response output *provided that, on the basis of the prior training procedure (involving intermittent primary reinforcement), the secondary reinforcer itself has been*

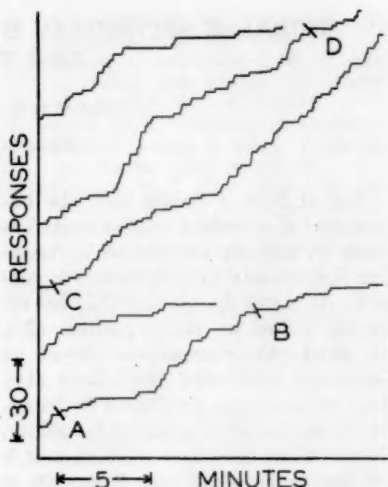


FIG. 1. Cumulative curves for acquisition and extinction of a bar-pressing response which received only secondary reinforcement. The two lower curves are for the animal's performance on Day 1; the three upper curves are for Day 2. From the origin to A, all responses were reinforced. From A to B, a one-minute fixed-interval schedule was used. Extinction began at B. At C (second day), the fixed-interval reinforcement schedule was resumed. Extinction began again at D. Each step rise in the curves represents one bar response.

*made "strong enough" to withstand a schedule of this kind.* If  $S^r$  is relatively weak, as in the usual sort of secondary reinforcement experiment, its effectiveness will wear out before any such schedule has a chance to operate.

Thus, two kinds of intermittent reinforcement are involved in the experimental procedure just described. One is intermittent reinforcement in the establishment of the secondary reinforcer, i.e., intermittent association of primary reinforcement with the secondary reinforcer. The other is intermittent administration of the secondary reinforcer itself, following the occurrence of the new response to be thereby strengthened. This second intermittent pro-

cedure may follow any of various schedules, typical results for one of which are shown.

The distinctive characteristics of the foregoing experimental procedure may be listed as follows:

1. The secondary reinforcing stimulus,  $S^r$ , is *discrete*. That is, it is not, as in many studies, some fixed characteristic of the experimental situation which persists throughout time—such as a continuous light in a Skinner box (4, 5), or the color of a runway goal box (16). In this respect it parallels a typical primary reinforcing event, such as the appearance (and disappearance) of a bit of food or water. The click of a food-delivery magazine, for a hungry rat, is a discrete event which may be presented anew following selected instances of a new response which is to be strengthened. Moreover, the stimulus thus used has an independently observable effect on behavior *before* the test procedure begins. The sound of magazine operation, following bar pressing (in an ordinary Skinner box), releases food-approaching behavior. In the same way the stimulus (buzzer) used in the present experiment leads to water-aperture approach during the training period, before its secondary reinforcing properties are independently demonstrated. In the case of something such as the color of a goal box, no such initial control of behavior is observed.

2. During training, water never appears without being preceded by  $S^r$ , but the appearance of  $S^r$  does not insure the appearance of water. In this respect the procedure is just the opposite of that used by Melching (12), where secondary reinforcement implies primary reinforcement but not the reverse.

3. The intermittent-reinforcement ratios used (both in the association of  $S^r$  with water and the presentation of  $S^r$  following the bar response) are relatively high. Most experimenters have

used ratios of the order of 1:2 or 1:3. In this respect the present procedure differs from that used by Saltzman (16) and Clayton (2), but is like that of Dinsmoor (4, 5). In order to achieve a high ratio of reinforcement, approximation by successive steps appears to be essential.

4. The reinforcing effect of  $S^r$  is clearly exhibited in the behavior of an *individual animal*. The observed experimental effect is so large that it is not necessary to use statistical criteria to demonstrate its reality. The same clear-cut effect has been obtained over and over again, with many animals.

## II. HOW DOES INTERMITTENT REINFORCEMENT INCREASE RESISTANCE TO EXTINCTION?

In the preceding section a method, admittedly rather complex, has been reported whereby highly stable and powerful secondary reinforcement can be demonstrated. This method is successful because it makes the phenomenon in question unusually resistant to extinction, and this end is achieved by the use of schedules of partial reinforcement. In order to get a better theoretical grasp of what is involved here, it will be useful to review a part of the history of the partial-reinforcement procedure.

The notion that failure or adversity, if ultimately followed by success, strengthens an individual's "character" and capacity to persevere is a very old one. The writer of the Book of James in the New Testament shows a clear appreciation of it. And Jenkins and Stanley, in their 1950 review, have this to say:

Partial reinforcement is such a ubiquitous phenomenon that one wonders why psychologists for so long failed to examine its consequences. . . . The animal trainer feeds his dogs only after several occurrences of the desired act, the clerk receives his check only once each month, the psychologist publishes his



findings only after prolonged endeavor, the gambler continues to bet despite an infrequent payoff, ad infinitum. Continuous reinforcement seems to be the exception, not the rule (10, p. 196).

When Skinner (19), first, and then Humphreys (9) and others reported experimental evidence on this score, their findings were more surprising theoretically than empirically. And the reason, as Jenkins and Stanley note, was this:

A straight S-R interpretation, that reward strengthens a response and omission of reward weakens it, would predict a weaker response with partial reinforcement than with 100% reinforcement in acquisition, performance, and extinction. The facts of post-partial extinction are clearly opposed to this position [in one important respect] (10, p. 222).

Numerous attempts have been made both to modify reinforcement theory so as to make it account for such findings and, by its opponents, to discredit the theory. No systematic attempt will be made here to cite the resulting literature. Instead, the writer will sketch what seems to be the currently most justifiable interpretation of this phenomenon.

Following related suggestions by Skinner (19) and Humphreys (9), Mowrer and Jones (15) in 1945 advanced what they termed the "discrimination hypothesis" to account for the effects of partial reinforcement upon resistance to extinction. To quote:

If, during acquisition, a response (conceived as a more or less isolated movement) occurs repeatedly but is rewarded only now and then, the transition from acquisition to extinction will not be discriminated as sharply as if acquisition has involved reward for each and every response. With "faith" thus established that failure will ultimately be followed by success, "discouragement" is slower to set in (ergo, greater "resistance to extinction") when there is a change in objective conditions from acquisition (occasional reward) to extinction (no reward whatever) . . . (15, p. 309).

Or, as formulated elsewhere, the notion would be that ". . . resistance to

extinction is greater following intermittent reinforcement because the subject has difficulty in discovering that the extinction procedure is extinction, rather than just a continuation of positive training" (14).

One may thus say that the greater resistance to extinction which occurs following partial reinforcement is due to the fact that the subject is "fooled" or "tricked" into misperceiving as a continuation of acquisition (ultimate reward likely) what is in reality extinction (no possibility of reward). This type of thinking has received experimental confirmation in a study reported by Bitterman, Feddersen, and Tyler (1). However, it is cast in a terminology that is frankly "cognitive" and needs, if possible, to be translated into more objective language. This, fortunately, is not difficult to do.

Let us begin by considering the following situation. A hungry rat is put into an experimental set-up in which each depression of a bar causes a pellet of food to appear in a nearby trough. After the bar-pressing response is well stabilized, with 100% reinforcement, the procedure is altered so that bar pressing suddenly ceases to produce food until a minute has elapsed. As a result of making several "extinction," i.e., unrewarded, responses, two things happen to the rat. Each unrewarded response leads to some decrement in response strength, but also—and for present purposes, more importantly—each unsuccessful response *becomes an event which precedes the reinforcement which is going to come at the end of the minute*. For this state of affairs to obtain, it should be noted, the response strength must be great enough that the response will continue to occur from time to time throughout the minute interval—that is to say, the shift from 100% reinforcement to less-than-100% reinforcement must not be so drastic that the response will completely

extinguish before the possibility of the next reinforcement is at hand.

Another way of expressing this position is as follows. During 100% reinforcement a number of stimulus events have been the occasion upon which food is received. These have included the stimuli arising from the bar and its vicinity, the proprioceptive consequences of mounting and beginning to press the bar, and, especially, the stimuli arising from eating the food pellet which was obtained on the just-previous press. During intermittent training, however, an additional element is added to this complex of stimuli which set the occasion for reinforcement—namely, *the stimulus consequences of a number of previous bar-presses which have not been reinforced*. It is not surprising, then, that this discrimination should continue to operate for some time into the extinction session. And, particularly, if *many* unreinforced bar presses have been experienced during acquisition, it is quite reasonable to suppose that the discriminative control would hold up *far* into the extinction session before breaking down.

This argument has been developed in greater detail by Skinner (20), who has shown that the form of the curves which are generated under different kinds of schedules, both during acquisition and during extinction, can be readily accounted for by such assumptions as these.

Another important implication of the discrimination hypothesis, which has not often been taken into consideration by theorists in dealing with partial reinforcement, should be stressed. The implication is this: *Behavior will be maintained without decrement under a particular intermittent schedule only if the change from 100% reinforcement to less-than-100% reinforcement is gradual enough to maintain discriminative control*. For example, if an animal is

switched from continuous reinforcement to a fixed ratio of 1:100, the response will extinguish. However, going upward in small steps—say from 1:1 to 1:10 to 1:20, etc., and eventually to 1:100, will maintain the performance.

The implications of this well-substantiated empirical finding are rather far-reaching, especially when considered in relation to the usual kind of formulations of partial reinforcement and of probability learning. It would seem that frequency of reinforcement is not simply a parameter which has an orderly relation to "habit strength" whenever it is introduced into an experimental session. Rather, the effects of given frequencies of reinforcement will depend critically upon the manner in which the intermittency is introduced into the session. At least this seems to be true for the free-responding situation.

In other words, no generalizations can be made about given frequencies of reinforcement, *per se*, without further specification of the initial and boundary conditions of the case. A curve showing "habit strength" as a function of frequency of reinforcement would be meaningless, since a given frequency might or might not maintain performance, depending upon the conditions under which it were approached. Introduction of a frequency, that is to say, is not "independent of path" in the mathematical sense.

### III. INTERMITTENCY AND SECONDARY REINFORCEMENT

As originally expressed by Hull, the principle of secondary reinforcement was simple and straightforward: A stimulus which occurs "consistently and repeatedly" in conjunction with a reinforcing state of affairs acquires, itself, reinforcing potential. As will become clear in the next section, the simplicity of this principle is elusive; and certain

amendments will be necessary in order to account for all the relevant data. Once again the paradigm of *discrimination* will prove useful.

Before pursuing this line of thinking, however, let us return to the secondary reinforcement procedure reported in the first section.

In secondary reinforcement, we have seen, there are two kinds of "conditioning" or "association" processes going on, namely: (a) association of the potential secondary reinforcer with primary reinforcement; and (b) subsequent presentation of this secondary reinforcer as a consequence of some selected response. Theorists have not always clearly distinguished between these two processes. Although drawing the distinction, Hull (8) stressed only the second and neglected the first. Whereas he offered much mathematical machinery in order to deal with the strength of S-R connections as a function of number of reinforcements and other variables, Hull said relatively little about the manner in which the strength of secondary reinforcers builds up and decays. Although a "repeated and consistent" association was held to be necessary, the form of the relationship here was not indicated with the same explicitness as in the case of other variables.<sup>3</sup>

<sup>3</sup> Another ambiguity, which may turn out to be of the highest importance, concerns the way in which the "strength" of a secondary reinforcer is to be measured. In one sense this "strength" might be presumed to vary along the same dimensions as does amount and quality of a reinforcing agent, that is, by leading to a *stronger habit* when presented a given number of times as a consequence of responding on the part of the animal. Another indication of the potency of the secondary reinforcer is the extent to which behavior can be established and maintained by its presentation in the absence of primary reinforcement, i.e., how long it lasts before wearing out. Subsequent analysis may, in fact, indicate that variation on this latter dimension may occur within wide limits, whereas there may be no variation

In the preceding section we have attempted to explain why intermittent reinforcement of *responses* results in greater resistance to extinction. There remains, however, the problem of dealing with intermittency in the association of S<sup>r</sup> with primary reinforcement (or in Hull's terms, a repeated but *inconsistent* association), which is something quite different.

An example from human behavior may help to clarify this point. When the telephone rings, we are not always reinforced for answering it. But when it rings, we predictably answer. This kind of intermittency of reinforcement is to be distinguished from that in which a given response occurs "spontaneously," over and over again, and is only occasionally reinforced. Whereas much is known experimentally about the effect of intermittency in the case of nondiscriminative, freely occurring behavior, the effects of intermittency in the case of discriminative behavior is only just beginning to be studied. As the research findings reported in this paper indicate, discriminative behavior, with intermittent reinforcement, can be maintained at high strength. But the upper limit of the ratio of reinforced to nonreinforced occasions which is here possible is probably not so high as in the case where the responses occur freely (ratios as high as 1/100 or 1/1000, and even higher, have been reported).

The intermittent secondary-reinforcement technique, as described above, is analogous. We are dealing with intermittent reinforcement of a response (running to the water aperture), but by the addition of a discriminative stimulus—the buzzer—we now have a situation in which the *experimenter can reliably and predictably control the response by virtue of his ability to manipu-*

at all on the first dimension. Discussion in later sections should clarify this point.

late the buzzer presentations. The event then may be made a consequence of prior behavior on the part of the animal (bar pressing), and the experimenter can observe the effect on the frequency of this behavior. By using the variable ratio procedure, stability of behavior following the buzzer is insured. Long runs of  $S^r$  occurrences with no water immediately forthcoming will have been encountered during training; and when extinction (a shift from occasional reinforcement to no reinforcement) is begun, the change will not be very great. This point will be discussed in greater detail below, but we may anticipate it by saying that this specific response—approaching the water aperture—may itself be a critical event in determining the reinforcing properties of  $S^r$ . Any procedure which insures that this response “lasts longer” (remains under the discriminative control of the signal) during the later test procedure will then augment the secondary reinforcing potential of the signal.

Granted that the secondary reinforcer will last for some time during the test session—that is to say, that it is “good for” a certain number of presentations following the desired new behavior—there is still the problem of maximizing the work output of the animal. It is here that the *second* intermittent procedure—the familiar intermittent reinforcement of a *response* (to which the discrimination hypothesis was originally designed to apply)—is employed. By presenting  $S^r$  intermittently during the test session, the magnitude of the reinforcing effect is increased in three ways. (a)  $S^r$  presentations are stretched out over a longer period of time, and more work is required of the animal for each of the available  $S^r$  presentations. (b)  $S^r$  presentations tend to be distributed, rather than massed, and for this reason also the reinforcing effect may diminish more slowly. (c) The resistance to ex-

tingtion of the new response is increased, so that when  $S^r$  is itself withdrawn, the animal may nevertheless continue to perform the new response for a considerable period of time thereafter.

We are thus dealing with a controlling event,  $S^r$ , which is diminishing in potency at the same time that it is being utilized. Each presentation of  $S^r$  strengthens some behavior, and also leads to a decrement in its effectiveness on the next presentation. The problem is to use  $S^r$  judiciously while it is effective. We might conceive of an experimenter with a supply of hypothetical food pellets which, by some magical means, grow smaller and smaller with each presentation to the animal, until they finally reach microscopic size, or disappear entirely. If the experimenter faces the problem of getting as much work out of the animal as possible, over as long a period of time as possible, an intermittent reinforcement procedure would clearly be in order.

#### IV. SECONDARY REINFORCEMENT AND DISCRIMINATION FURTHER CONSIDERED (THE DINSMOOR HYPOTHESIS)

Following suggestions previously put forward by Schoenfeld, Antonitis, and Bersh (17), Dinsmoor (3) in 1950 advanced the thesis that secondary reinforcement and discrimination, in some kinds of situations, are identical—or, at least, that any stimulus which is a secondary reinforcer also has cue properties, and vice versa. The great importance of this thesis—and Dinsmoor reported experimental results which substantially confirmed it—is only just beginning to be appreciated. One of the major implications is this: A stimulus which is “repeatedly and consistently” paired with primary reinforcement presumably will *not* acquire secondary reinforcing properties unless, in addition, it has first played a discriminative role

of some kind, as a result of a differential reinforcement procedure.

On the basis of his experiment, Dinsmoor concludes:

It no longer seems to be necessary to provide separate theoretical accounts of the manner in which a discriminative or reinforcing stimulus gains or loses the power to raise the rate of a response. The previous distinction between the two types of stimulus appears to be reduced to a distinction between two categories of temporal schedule for the administration of the stimulus. This distinction is relevant chiefly to the determination of the distribution of the stimulus effects among the variety of responses in the organism's repertoire (3, p. 471).

What Dinsmoor means by the latter remark, apparently, is this: If a stimulus appears before a response which receives primary reinforcement (sets the occasion for a response), we call that stimulus a *cue* ( $S^D$ ); whereas, if the same stimulus is produced, not by the experimenter, but by some distinctive action on the part of the subject, we refer to it as a *secondary reinforcer* ( $S^r$ ) because it increases the likelihood of recurrence of this response. Said otherwise, a secondary reinforcer is a cue which the subject himself provides, by means of his own behavior. And when a stimulus gains or loses one capacity, it at the same time gains or loses the other.

The question now naturally arises as to how these considerations bear upon the so-called "discrimination" hypothesis previously discussed, in Section II. The relationship seems to be straightforward and consistent. According to the Dinsmoor thesis, a secondary reinforcer is a stimulus which enables the subject to discriminate between the conditions ( $S^A$ ) under which a particular response, if it occurs, will *not* be reinforced and conditions ( $S^D$ ) under which this response *will* be reinforced. And the secondary reinforcement capacity of such a stimulus is shown by the fact

that the subject will learn to make a *new* response if that response produces the cue stimulus,  $S^D$ .

Now, in a partial-reinforcement training procedure, a special condition is introduced. Here  $S^A$  (absence of  $S^D$ ) still means that the response in question will not "work"; i.e., in the situation described in Section I, the rat never gets water as a result of going to the water aperture when the buzzer has not sounded. But there is this difference: although the buzzer is thus a cue—which must be responded to (lest water be available, but not discovered, as it were)—it is an inconsistent one. As has been mentioned above, this inconsistency of reinforcement does not diminish the discriminative control which is exerted by the stimulus. And because it has been thus intermittently reinforced during acquisition, the lack of reinforcement is, so to say, no novelty when it occurs during extinction.

#### V. IS A SECONDARY REINFORCER MORE EFFECTIVE IF IT "RELEASES" BEHAVIOR?

The Dinsmoor hypothesis points to an equivalence between two functions of a stimulus which were formerly held to be separate. Although the results here are not as yet unequivocal, enough evidence now exists to justify the use of a discrimination training procedure whenever an effort is being made to give a stimulus secondary reinforcing properties. One may note these results, however, and still ask the question, Why? We must now examine briefly the possibility that the equivalence between discriminative stimuli and secondary reinforcing stimuli may be derived from more basic principles. Three possibilities suggest themselves.

1. In watching the behavior of the animals during the procedure described in Section I, one is reminded of Guthrie's



lastness-of-response hypothesis (7), or the postremity principle, as Voeks (24) has called it. Certainly a striking thing about the buzzer is the promptness and reliability with which it calls the subject away from whatever else it is doing and causes it to go to the water aperture. And when the bar is made available and so connected that the subject, by depressing it, can produce the buzzer (at least intermittently), one might argue that the response is preserved, or "protected." That is, the buzzer guarantees that the *last* response made will be one of bar pressing, since the buzzer then comes on and "takes the subject out of the situation."

A similar interpretation of the effects of reinforcers (both primary and secondary) has been suggested by Verplanck (23), and Estes (6). Following this kind of thinking in connection with *secondary* reinforcement, it is easy to see why the Dinsmoor restriction—that  $S^r$  must also be  $S^D$ —would be essential. When, following training, the stimulus is presented as a consequence of a response, it will elicit the behavior which it has formerly controlled in its discriminative role—in the present context, approaching the magazine. This uniform behavior, as long as it lasts, will "take the animal out of the situation," just as does a primary reinforcer. The Dinsmoor hypothesis may, in the light of this interpretation, be restated in somewhat stronger form: *A stimulus will be effective as a secondary reinforcer for new behavior if and only if it has some response already conditioned to it.* The secondary reinforcer, that is to say, must predictably "release" some behavior as a result of a prior training procedure; and when it then is used to strengthen new behavior, it is this response which is "released," not the stimulus *qua* stimulus which is the critical event. The question "How can secondary reinforcers be made to last longer?" may then be rephrased,

"How can discriminative stimuli be made to retain control over uniform behavior longer?"

Whether such specific control is both *necessary* and *sufficient* for reinforcement is a question which only further research can answer. This position would seem to be demanded by strict contiguity theorists such as Guthrie.

2. The former, weaker hypothesis—that discrimination is necessary, but not sufficient—could be readily assimilated by a variety of current theories of learning. A position such as that of Spence (22) or Seward (18), for example, could hold that the effect of a stable magazine-approach response depends upon  $r_g$ 's which are consistently elicited and protected from extinguishing.

3. Yet another possibility suggests itself. Because the buzzer reliably elicits running to the water aperture, the proprioceptive and other feedback from this response also become effective and the *range* of stimulation received is thereby enlarged. Hence, when the subject is later given an opportunity to press the bar, it is plausible to hold that there is now, so to say, a *double* course of secondary reinforcement: the buzzer (immediate feedback from bar pressing) and the stimulation produced by the behavior which the buzzer releases (mediated feedback from bar pressing). Thus, to generalize, one might conclude that any procedure where  $S^r$  is also an  $S^D$  will result in more potent secondary reinforcement because here there are more varied and more numerous stimuli provided by the mediating behavior. Perhaps much the same effect could be achieved by having  $S^r$  consist, not of a single, segmental stimulus change, but of several such stimuli. The latter might be more properly termed a *change in the situation*, thus being more nearly comparable to the change brought about by the animal's own response to  $S^r$  and the resulting stimulus consequences. This

is, however, at best conjectural; and the whole question of precisely why a response to an S<sup>r</sup> makes it so much more efficient as a secondary reinforcer is still far from settled.

#### VI. SUMMARY AND CONCLUDING REMARKS

Mowrer (13) has recently presented a theoretical interpretation of learning which depends critically upon the principle of secondary reinforcement. This theory is, in essence, a searching analysis and reformulation of the concept of *habit*, which is so ubiquitous in contemporary learning theory. Secondary reinforcement, as an experimental paradigm, is used as a major explanatory principle.

In the past, one of the disconcerting things about secondary reinforcement has been the extremely transitory and unstable nature of the demonstrated laboratory effect. One thing seems clear: a theory which leans so heavily upon this principle is doomed at the outset if the raw empirical phenomenon cannot be demonstrated in more potent form. Before using the paradigm as an explanatory device to account for (reduce to more basic principles) the most complex forms of mammalian behavior, there seems to be the anterior question as to whether, in fact, secondary reinforcement is *real*, *qua* phenomenon. Although experimental work within the last three decades has provided suggestive evidence, the results have been far from unequivocal. Actual experimental demonstration—such that a disinterested observer could view the animal in question and be convinced that “something is really happening”—has been lacking.

It has been the purpose of this paper to discuss a method by which secondary reinforcement can be made highly effective and stable, thus giving the needed empirical buttressing to the kind of

theorizing cited. This method depends upon the use of *intermittent* reinforcement, thereby also further accenting the very great importance of the intermittent reinforcement technique for psychological theory.

#### REFERENCES

1. BITTERMAN, M. E., FEDDERSON, W. E., & TYLER, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, 66, 456-464.
2. CLAYTON, F. L. Secondary reinforcement as a function of reinforcement scheduling. *Psychol. Reports*, 1956, 2, 377-380.
3. DINSMOOR, J. A. A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *J. exp. Psychol.*, 1950, 40, 458-472.
4. DINSMOOR, J. A. The effect of periodic reinforcement of bar pressing in the presence of a discriminative stimulus. *J. comp. physiol. Psychol.*, 1951, 44, 354-361.
5. DINSMOOR, J. A. Resistance to extinction following periodic reinforcement in the presence of a discriminative stimulus. *J. comp. physiol. Psychol.*, 1952, 45, 31-35.
6. ESTES, W. K. Generalization of secondary reinforcement from the primary drive. *J. comp. physiol. Psychol.*, 1949, 42, 286-295.
7. GUTHRIE, E. R. *The psychology of learning*. (Rev. ed.) New York: Harper, 1952.
8. HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
9. HUMPHREYS, L. J. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, 25, 141-158.
10. JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, 47, 193-234.
11. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
12. MELCHING, W. H. The acquired reward value of an intermittently presented neu-

- tral stimulus. *J. comp. physiol. Psychol.*, 1954, 47, 370-374.
13. MOWRER, O. H. Two-factor learning theory reconsidered, with special reference to secondary reinforcement and the concept of habit. *Psychol. Rev.*, 1956, 63, 114-128.
  14. MOWRER, O. H. Learning theory and behavior. Urbana, Ill.: Author, 1957.
  15. MOWRER, O. H., & JONES, H. M. Habit strength as a function of pattern of reinforcement. *J. exp. Psychol.*, 1945, 35, 293-311.
  16. SALTZMAN, I. J. Maze learning in the absence of primary reinforcement: a study of secondary reinforcement. *J. comp. physiol. Psychol.*, 1949, 42, 161-173.
  17. SCHOENFELD, W. N., ANTONITIS, J. J., & BERSH, P. J. A preliminary study of training conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1950, 40, 40-45.
  18. SEWARD, J. P. How are motives learned? *Psychol. Rev.*, 1953, 60, 99-110.
  19. SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century, 1938.
  20. SKINNER, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, 57, 193-216.
  21. SPENCE, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, 54, 1-8.
  22. SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.
  23. VERPLANCK, W. S. The operant conditioning of human motor behavior. *Psychol. Bull.*, 1956, 53, 70-83.
  24. VOEKS, V. W. Formalization and clarification of a theory of learning. *J. Psychol.*, 1950, 30, 341-362.

(Received February 25, 1957)

## AN OPPONENT-PROCESS THEORY OF COLOR VISION

LEO M. HURVICH<sup>1</sup> AND DOROTHEA JAMESON<sup>1</sup>

*Eastman Kodak Company*

The two major theoretical accounts of color vision are those classified as the Young-Helmholtz and the Hering types of theories. For many years the former has been judged by most workers in the field to provide the simplest explanation of the way in which light stimuli give rise to color sensations. The advantages that appear to favor the Young-Helmholtz three-component hypothesis are two: it is parsimonious, and its postulates are easily quantifiable and hence subject to precise experimental test. In its parsimonious and easily quantifiable form, the theory is simple: in addition to the rods which subserve twilight vision, the eye contains three kinds of cone photoreceptors; each type of cone contains a differently selective photochemical substance; each is associated with its own specific nerve fiber; and each cone-photochemical-nerve fiber system is correlated with one of the three specific "fundamental" color sensations, namely, red, green, and blue (or violet). All sensations are considered as compounded of varying amounts of these three excitatory systems, with white arising from equal and simultaneous excitation of all three, and yellow from equal red and green excitations.

The Young-Helmholtz three-cone, three-nerve, three-sensation theory derives directly from the basic fact of color mixture, namely, that all visible hues can be matched by the mixture, in proper proportions, of only three physical light stimuli. Based squarely on this fact, the theory is readily quantified in terms of the three measurable variables of color mixture experiments. But the three measured variables, it

must be emphasized, are the three physical light stimuli used in the color mixture experiments; they are not the postulated three "fundamental" color sensations, for with each different stimulus triad used for color matching a different and equally valid triad of color mixture functions is obtained. Consequently, throughout some hundred years since the original formulation of the idea, a continued series of attempts has been made to find the proper transformation of the three measured color-mixture curves that will bridge the gap and yield the unique spectral distribution curves of the desired physiological correlates of the three postulated "fundamental" color sensations. An infinity of such transformations is available for trial, and almost every serious adherent of the theory has proposed at least one new set of "fundamental sensation curves" (48, pp. 368-372). The search, however, continues, because serious defects have been found in every proposal made thus far. When the explanatory or predictive power of the theory in any given quantified form is tested it cannot handle more than a limited number of facts satisfactorily (11, p. 805).

Moreover, some facts of color experience seem unassimilable into the framework of the simple Young-Helmholtz theory with its three independent, fundamental, process-sensation systems. How can this system of three independent processes be made to account, for example, for the apparent linkages that seem to occur between specific pairs of colors as either the stimulus conditions or the conditions of the human observer are varied? Why should the red and green hues in the spectrum predominate

<sup>1</sup> Now at New York University.

at low stimulus levels, and the yellow and blue hue components increase concomitantly as the spectrum is increased in luminance (43)? Why, as stimulus size is greatly decreased, should discrimination between yellow and blue hues become progressively worse than that between red and green (4, 10)? Why should the hues drop out in pairs in instances of congenital color defect, or when the visual system is impaired by disease (29, 31)? On the other hand, since the sensation of white is granted no special physiological process in this parsimonious theory, but occurs as the fusion product of three equally large fundamental hue sensations, how account for the large degree of independence of white and chromatic qualities when the adaptation of the visual system is varied (37, 41)?

As more and more *ad hoc* hypotheses are added to the original Young-Helmholtz formulation in order to answer these and other problems forced by the increasing accumulation of experimental data, we naturally find the formulation becoming less and less precise and quantifiable, and obviously less parsimonious. We also find, however, that exactly those phenomena that require modification and extension of the simple "three-color theory" remind us more and more of its chief theoretical rival, the Hering theory of three paired, opponent color processes.

In view of this situation, it seems highly desirable that we take a close second look at Hering's alternative approach to an understanding of color phenomena. The vast accumulation of psychophysical data for which any adequate theoretical proposal must account requires that the basic postulates of the theory, as outlined qualitatively by Hering (13, 14), be restated in quantitative terms for such a critical scrutiny to be most meaningful. This paper will review our attempt to provide such a

quantitative restatement, and will summarize briefly some of the critical comparisons between the theoretical deductions and relevant psychophysical data. (Detailed quantitative accounts are given in 21, 22, 23, 25, 26, 27.)

#### BASIC SCHEMA FOR THE HERING THEORY

##### *The Three Variables*

The Hering theory is like the Young-Helmholtz theory in that it, too, postulates three independent variables as the basis for color vision, but the Hering variables are three pairs of visual processes directly associated with three pairs of unique sensory qualities. The two members of each pair are opponent, both in terms of the opposite nature of the assumed physiological processes and in terms of the mutually exclusive sensory qualities. These paired and opponent visual qualities are yellow-blue, red-green, and white-black.

The basic schema for the opponent-colors mechanism is shown diagrammatically in Fig. 1. The three paired opponent response systems are labeled *y-b*, *r-g*, and *w-bk*. The convention of positive and negative signs is used to

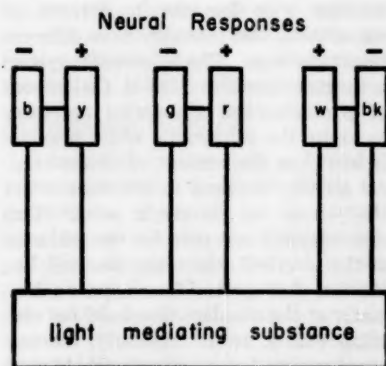


FIG. 1. Basic schema for Hering opponent-colors mechanism.



indicate that each neural system is capable of two modes of response that are physiologically opponent in nature, and that the paired sensory qualities correlated with these opposed modes of response are also mutually opponent or exclusive. That is, we may experience red-blues or green-blues but never yellow-blues, and we see yellow-greens or blue-greens, but never red-greens, and so on. In the absence of any external visual stimulus, the state of the visual system is assumed to be a condition of active equilibrium, and this equilibrium condition is associated with the neutral, homogeneous "gray" sensation perceived after a long stay in complete darkness. This sensation is quite different from the black experience of the white-black opponent pair. Blackness arises neither by direct light stimulation nor in the simple absence of light, but rather by way of either simultaneous or successive contrast during, or following, light stimulation of some part of the retina.

#### *Properties of Paired Systems*

The three pairs of visual response processes are independent of each other; that is, they have different response thresholds, they follow different laws of increase with increase in strength of stimulation, and probably have different time constants. The achromatic system is the most sensitive; that is, the amount of photochemical absorption necessary to excite the achromatic white response is less than the amount of photochemical activity required to stimulate either the  $y-b$  or  $r-g$  chromatic pairs. This characteristic accounts for the existence of the so-called achromatic interval, i.e., the fact that spectral lines appear achromatic at the absolute threshold for visibility (42, p. 167). Similarly, the red-green system has a lower threshold than the yellow-blue one. The failure of the yellow-blue system to respond at near-

threshold levels that are sufficient to activate the red-green system exhibits itself in the facts of so-called "small field dichromasy," in which the eye behaves, with respect to stimuli that are very small in area as well as of low intensity, in a manner similar to the congenital tritanope, i.e., a specific type of "color blind" individual for whom yellow and blue discriminations are impossible and the only hues seen are reds and greens (4, 49).

With increase in level of stimulation the different paired systems also show differences in rate of response increase, such that the achromatic response increase is probably the most rapid of the three, with the result that at very high intensities all spectral stimuli show a strong whitening, or desaturation, relative to their appearance at some intermediate luminance level (42, p. 168). Of the two chromatic pairs, the yellow-blue system, although exhibiting a higher threshold, shows a more rapid rate of increase in response with increase in luminance than does the red-green system. Thus, the mixed hues of the spectrum—the violets, blue-greens, yellow-greens, and the oranges—all vary systematically with increase in spectral luminance, and all show a tendency to be more blue or yellow, respectively, at high luminances, and more red or green at the lower luminance levels (the Bezold-Brücke hue shift phenomenon).

The opponent systems show a tendency toward restoring the balanced equilibrium condition associated with the neutral "gray" sensation. Thus excitation, say, of the  $r$  process in the  $r-g$  system results in a decrease with time in  $r$  responsiveness, and in an increase in the responsiveness of the opponent  $g$  process. If we think of the  $r$  process as perhaps associated with the building up of an electrical potential in the neural system, and of the  $g$  process as associated with the collapse of the po-

tential during impulse firing, then it is easy to see that as the neural potential is increased to higher values there will be a tendency to resist further build up, and also an increased disposition of the tissue toward impulse firing in order to restore the potential to its normal equilibrium value. Although we are not at all ready to ascribe a specific neural correlate of this sort to the postulated opponent processes at this time, the neurophysiological parallels are useful for conceptualizing the opponent-process notion as a real biological phenomenon.

To return to our example, if the responsiveness of the opponent *g* process tends to increase as *r* excitation is continued, then when the *r* stimulus is removed we can expect *g* activity to be released, strongly at first, then more slowly, and ultimately fading out as equilibrium is again approached. The sensory correlate of this reversal of opponent activities with removal of stimulation is, of course, the familiar phenomenon of the complementary after-image. If the stimulus (of constant magnitude) is not removed but continues to act for a considerable length of time, then the *r* process, whose responsiveness is being continuously decreased, will eventually cease to respond further, and a new equilibrium state will be reached. The disappearance of a sensory response with continued constant stimulation can be observed either by the *Ganzfeld* technique, in which the whole retina is uniformly illuminated by diffuse light (18), or by the "painted image" technique, in which optical means are used to fix a well defined image on the retina in such a way that its retinal position remains constant and independent of eye movements (39). By either method the eventual result of continued steady stimulation is a disappearance of the visual experience: the light seems to have gone out in the

*Ganzfeld* situation, or, in the fixed-image situation, the perceived object simply fades out of view.

Not only are the visual responses modified by changes in time in the excitabilities of the opponent processes, but they are also importantly affected by spatial interaction among the various elements of the visual field. Within certain limits there is evidence of summation of similar kinds of activity in adjacent elements, as in threshold responses for small stimulus areas (5, pp. 846-852). But perhaps more important for the over-all functioning of the visual system are the antagonistic interactions, such that *r* activity in one area induces *g* activity in adjacent areas, and similarly for the yellow-blue and white-black paired response systems. These opponent spatial induction effects are evident in all the familiar color and brightness contrast phenomena (35, pp. 138-142). They are probably also primarily responsible for the great visual-image clarity that characterizes vision in spite of the fact that the optical system of the eye is obviously imperfect, and that consequently the light image formed on the retinal surface lacks sharply defined boundaries (17, pp. 151-159). The spatial interaction causing intensification of opponent qualities at adjacent surfaces would seem an ideal sharpening device to sharpen up the initially blurred retinal image.

#### *Photochemical Postulates*

In addition to the various temporal and spatial induction effects, which are assumed to be based in the neural visual-response tissue, visual adaptation probably also involves changes in the photochemical activities that initiate the neural responses, since a certain amount of photochemical bleaching is expected to occur with continued exposure of the photosensitive materials to a retinal light stimulus. In order

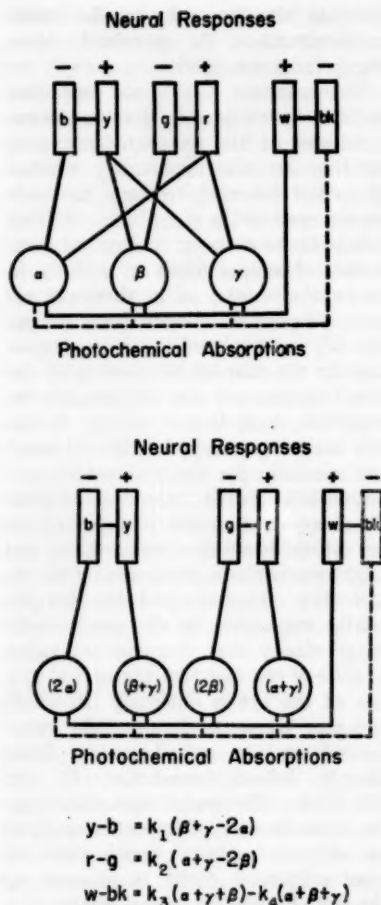


FIG. 2. Schematic diagram showing relations between photosensitive materials  $\alpha$ ,  $\beta$ , and  $\gamma$  and neural response processes  $y-b$ ,  $r-g$ , and  $w-bk$ .

for the three paired opponent-response systems to be selectively stimulated, there must, of course, be more than one substance available for photochemical mediation between the incident light and the neural excitation. Whatever the specific nature of the photosensitive materials, they must form a link in the system of three independent variables,

and hence we have postulated three independent photosensitive materials, which we may call  $\alpha$ ,  $\beta$  and  $\gamma$ .

Our schematic model now takes the form shown in Fig. 2A or 2B. The three independent photosensitive materials may be contained in discrete retinal units with complex interconnections to the neural response systems, as shown in Fig. 2A, or two or more of these materials may be combined in receptor units having simpler connections to the neural response systems, as diagrammed in Fig. 2B. There is no way of differentiating these models in terms of visual behavior; and however the three photochemicals may be segregated or combined in the retina, and whatever the number of different photoreceptor units, there remain only three independent photosensitive materials, and the theory remains a three-variable, opponent-colors schema.

#### QUANTIFICATION OF OPPONENTS THEORY

Since our aim is to present this schema in quantitative terms, one of the first questions that has to be asked is this: Is it possible to obtain by psychophysical experiment direct measurements of the spectral distributions of the three basic response variables of the Hering theory?

#### Measures of Achromatic and Chromatic Responses

It can fairly be assumed that the achromatic, white response is closely connected with the distribution of the brightness quality throughout the visible spectrum, and Fig. 3 therefore shows two functions (which we have measured by a threshold technique) that give the whiteness distribution of an equal energy spectrum for two observers (20). The induced rather than directly stimulated black component of the achro-

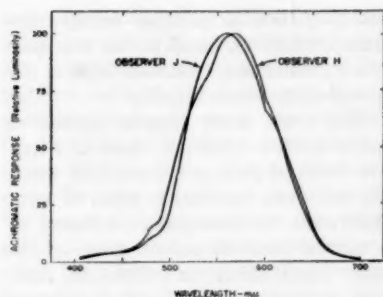


FIG. 3. Whiteness distribution of an equal energy spectrum for two observers.

matic white-black response pair has this same distribution, but of opposite sign, since the strength of the black contrast response is directly related to the magnitude of either the surrounding or the preceding whiteness or brightness.

A method for determining the spectral distributions of the paired chromatic responses is implicit in the opponents theory itself. Since the two members of each hue pair are mutually opponent or exclusive, then a yellow response of given strength should be exactly canceled by a stimulus that, taken alone, elicits the same magnitude of blue response, and a similar relation should hold between red and green responses. Thus a null method, based on the antagonism of the two members of each hue pair, can be used to measure the spectral distributions of the chromatic responses. In brief, a wave length is first selected that evokes, say, a blue hue response. The observer then views, in turn, a series of spectral wave lengths that appear yellowish in hue (yellow-greens, yellow, and yellow-reds). To each of these yellow stimuli just enough of the previously selected blue stimulus is then added exactly to cancel the yellow hue without introducing any blueness. The observer simply reports when the test field appears neither yellow nor blue; the hue remainder that he sees

may be green, neutral, or red, depending on the test wave length. Knowing the energies of the series of spectral yellow stimuli, and having determined experimentally the energy of the blue stimulus of fixed wave length that is required for the hue cancellation in each case, we can now plot the distribution of the relative magnitudes of yellow hue response evoked by the various test wave lengths. The procedure is simply reversed to obtain the distribution of

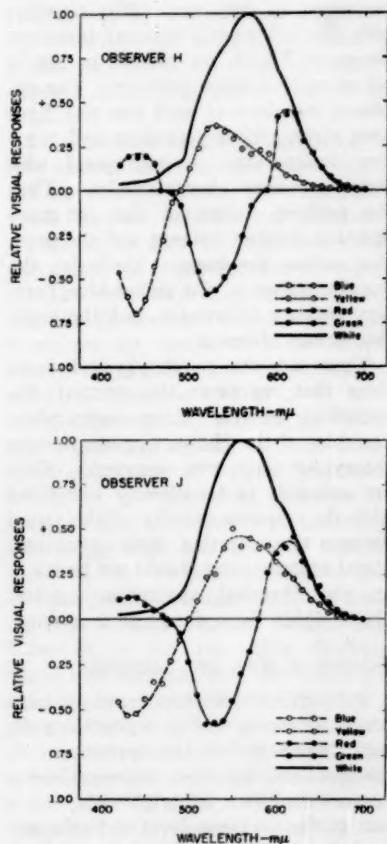


FIG. 4A and 4B. Chromatic and achromatic response functions for equal energy spectrum for two observers.

the blue component of the yellow-blue pair; that is, varying amounts of a fixed wave length of yellow hue are used to cancel the blue hue quality of a series of "blue" test wave lengths. By using a red stimulus of fixed wave length and variable energy to cancel the greens, and a green stimulus to cancel the reds, the spectral distribution of the red-green pair of chromatic responses is similarly determined.

Two sets of paired chromatic response vs. wave length functions that were measured in this way (25), together with the achromatic spectral functions shown in Fig. 3, are plotted in Fig. 4 for an equal energy spectrum. The opponent members of each hue pair have been given arbitrary positive and negative designations, to correspond with their opponent characteristics. Thus the positive values of the red-green function indicate redness, and the negative values greenness. Similarly, the positive values of the yellow-blue function indicate yellowness, and the negative values blueness.

These are the psychophysical functions that represent the spectral distributions of the three independent variables of the Hering opponent-colors theory for these two observers. They are assumed to be directly correlated with the response activity of the visual nervous tissue (retina, optic nerve, and visual centers), and should not be taken as photochemical absorption spectra, about which these data tell us nothing.

#### *Brightness, Hue, and Saturation*

The psychophysical opponent-response functions shown in Fig. 4 provide a direct description of the appearance of the spectrum, for these observers, for a neutral condition of bright adaptation and at the moderate level of luminance for which the functions were obtained. Thus, all wave lengths evoke some whiteness as well as hue; the whiteness

and brightness of an equal energy spectrum is relatively small at the two spectral extremes and relatively high at the intermediate wave lengths.

The short wave lengths appear as red-blue hues (violets); there is a narrow band of pure or unique blue where the red-green function is equal to zero; then come the blue-greens, followed by a narrow band of unique green at the wave length where the yellow-blue function is equal to zero; this is followed by the yellow-greens, and then pure yellow occurs at the second intersection of the red-green function with the zero ordinate value; and finally the yellow-red hues appear in the long wave length region (19). A quantitative expression for hue, a "hue coefficient," can be obtained by taking the value of one of the chromatic responses, say, the yellow value at 550  $m\mu$ , relative to the total of all chromatic responses at that wave length, in this case, yellow plus green.

The saturation of the color depends on the relative amounts of chromatic and achromatic responses. At the two spectral extremes where the chromatic responses are large relative to the white response, the spectral saturation is high. Where the reverse is true, spectral saturation is low. This can be expressed quantitatively in the form of a "saturation coefficient." To use the same example, the total of the yellow-plus-green values relative to the white plus yellow plus green is relatively low at 550  $m\mu$ , and this wave length appears much less saturated than does, say, either 440  $m\mu$  or 670  $m\mu$ .

#### *Color Mixture*

Since color-mixture experiments simply involve matching the three perceived qualities evoked by one stimulus by the proper mixture of three other stimuli, it is possible to determine the color-mixture relations that are inherent in the response curves of Fig. 4 for



any three arbitrarily selected mixture primaries. That is, the red-green value, the yellow-blue value and the white value of the total visual response to any wave length of unit energy are matched by the totals of the three corresponding values for the three mixture primaries when the latter stimuli are combined in the proper ratios. On paper, the color equations for most spectral matches require the admission of negative values for one of the mixture primaries. In actual color-mixture experiments, these negative values are realized by removing one of the mixture primaries from the matching field and adding it to the test stimulus.

To calculate, for example, the amounts of energy required for a color match to a given wave length  $\lambda$  by the mixture of the spectral primaries 460 m $\mu$ , 530 m $\mu$ , and 650 m $\mu$ , let  $a$  = the energy at 460 m $\mu$ ,  $b$  = the energy at 530 m $\mu$ , and  $c$  = the energy at 650 m $\mu$ . The three equations to be solved for these three unknowns  $a$ ,  $b$ , and  $c$  are then:

$$a(r_{460}) + b(r_{530}) + c(r_{650}) = r_{\lambda}$$

$$a(y_{460}) + b(y_{530}) + c(y_{650}) = y_{\lambda}$$

$$a(w_{460}) + b(w_{530}) + c(w_{650}) = w_{\lambda}$$

The values for  $r$  (or for  $-r$  when the response function is negative, indicating that the hue is green rather than red), for  $y$  (or for  $-y$  when the response is blue rather than yellow), and for  $w$  are then read from the response functions for unit energy for each wave length in question. (See Fig. 4.) The values  $r_{\lambda}$ ,  $y_{\lambda}$  and  $w_{\lambda}$  represent the unit energy response values for any spectral wave length for which a color-mixture equation is to be calculated. Solving this set of three equations for the three unknowns  $a$ ,  $b$ , and  $c$ , we then have a color-mixture equation of the form

$$a_{460} + b_{530} + c_{650} = 1_{\lambda}$$

This equation, which is expressed in energy units, may be converted to photo-

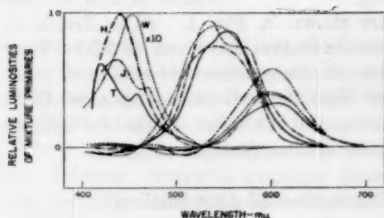


FIG. 5. Calculated color mixture functions for observers H and J and experimental color mixture functions for observers T and W (24, 47).

metric units in the usual way by multiplying each energy value by the relative luminosity (given by the achromatic response function) at the given wave length.

Color-mixture relations calculated in this manner for wave lengths  $\lambda$  from 420 m $\mu$  through 700 m $\mu$  from smoothed visual response data for two observers are shown in Fig. 5. The two additional sets of color-mixture functions (for the same three mixture primaries) that are shown for comparison in the figure are the results of actual color-mixture experiments by W. D. Wright and L. C. Thomson (24, 47).

Since the relations between the measured response functions and the color-mixture data are, as we have just seen, known for two individual observers, it is now also possible (by assuming specific spectral loci for the unique hues) to reverse the procedure and derive opponent-response functions from the color-mixture data for Wright and Thomson, or for any other observer whose color-mixture data are available. Since it seems preferable to develop a general theoretical model on the basis of a representative average, rather than a small number of individual observers, we have used for the model chromatic and achromatic response functions derived from the average color-mixture data for the CIE international standard observer (30). These derived functions

are shown in Fig. 6. (The details of the derivation are given in 22.) They are, of course, smoother and more regular than the individual, measured functions, but in other respects they are quite similar (compare Fig. 4).

### Photochemical Distributions

The specific set of  $\alpha$ ,  $\beta$ , and  $\gamma$  photosensitive absorption functions that have been assumed for the theoretical model are shown in Fig. 7. These curves have not been measured, and they have the particular forms shown and the high degree of overlap exhibited because of the specific interrelations that we have postulated a priori between the photochemical events and the neural response activities of the visual opponent mechanisms. Once the photopigments actually present in the human retina have been identified by the biochemists, the visual theorist will have no need to make such a priori postulates, and the specific interrelations required between the identified photosensitive materials and the neural processes underlying the color responses can easily be deduced.

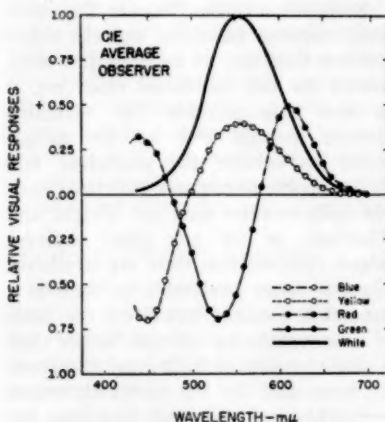


FIG. 6. Theoretical chromatic and achromatic response functions for equal energy spectrum for CIE average observer.

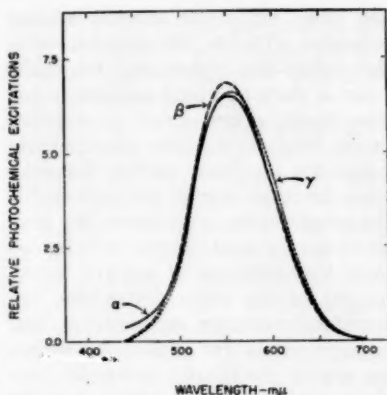


FIG. 7. Spectral distribution curves for assumed photosensitive materials.

As matters now stand, however, the functions shown in Fig. 7 meet the basic demands of the known facts, and any changes in these theoretical absorption functions that will no doubt be required by results of photochemical researches will not importantly affect any of the basic postulates of the theoretical model. The broadness and similarity of shape of all three selective functions that we have assumed are characteristic of all visual pigments so far identified in a variety of animal species (2).

These assumed photopigment distributions do not enter into the consideration of color phenomena, in normal vision, until we come to an examination of some of the phenomena of chromatic adaptation in which selective photochemical bleaching seems to act as one of the important determinants. The other determinants are, of course, the spatial and temporal induction effects in the neural opponent response processes that have been mentioned earlier.

### Dependence of Hue and Saturation on Both Wave Length and Luminance

What are the visual phenomena for which the model can account? As we

have already indicated, the measured chromatic and achromatic response functions provide a direct and quantifiable description of the color sensations evoked by any stimulus of specified wave-length composition (23). The achromatic, white function is taken as a direct expression of spectral brightness. Spectral hue, which is determined by the chromatic responses evoked by each wave length, can also be expressed quantitatively as a coefficient value relating the magnitude of response of one chromatic system to the total of all chromatic responses at that wave length. An example of such a hue coefficient function for a moderate level of luminance is shown in Fig. 8. It is clear, from the varying rate of change in the hue coefficient function from one region of the spectrum to the next, that an observer's ability to discriminate among neighboring wave lengths on the basis of hue changes alone will also differ for the different regions of the spectrum. This discriminative capacity is obviously also quantifiable in terms of the amount of wave-length change required to elicit a threshold change of fixed amount in the value of the hue coefficient. With change in the luminance at which the spectrum is presented, these coefficient

functions will be altered, in the sense that the yellow-blue values will increase at the higher luminances, and will be diminished at the lower luminances. This is so because, in accordance with the different energy-vs.-response function postulated for the yellow-blue system as compared with the red-green one, as the excitation level is increased, the yellow and blue spectral responses will be uniformly magnified relative to the red and green ones at the higher levels, and uniformly diminished at the lower levels. Although the exact differential between the two paired systems is not known, under certain circumstances an over-all difference in response magnitudes of approximately 20 per cent seems to occur for a log unit change in luminance. Thus, at some wave length for which, say, the red and yellow responses are equal at a luminance of 10 mL, the yellow will be about 20% greater than the red at 100 mL, and about 20% less at a luminance of only 1 mL. If we assume this 20% differential between  $y-b$  and  $r-g$  response magnitudes per log unit of luminance change as a reasonable value, and compute the spectral hue coefficients for a range of approximately three log units of luminance variation, then we can specify the amount of hue shift associated with a change in intensity of any wave length. Conversely, we can also specify the wave length changes necessary to maintain a constant hue sensation (constant hue coefficient value) as the luminance is increased or decreased. The latter procedure has been used to obtain the functions shown in Fig. 9, and the curves in the upper part of the figure are functions measured by Purdy in an actual experiment of this sort (38).

These hue phenomena do not involve the achromatic response pair at all, and depend only on the two paired chromatic response systems. Whatever the

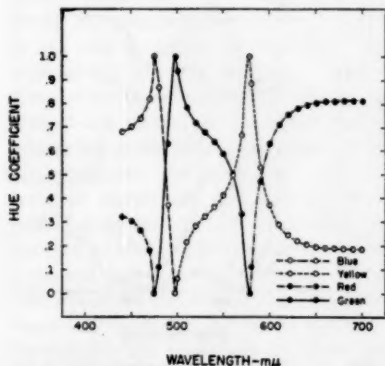


FIG. 8. Spectral hue coefficients. Moderate luminance.

chromatic response to a given stimulus, the perceived color saturation clearly will also depend on the extent to which the white system is simultaneously responding. For any given amount of chromatic response, the color will obviously appear less saturated if there is a large magnitude of white response to dilute the color, and more saturated if the white component of the total response is relatively small. The perceived saturation of the spectrum is also expressed as a quantitative coefficient function. (See Fig. 10.) Here the value taken as the saturation coefficient is the ratio of the total chromatic to the chromatic-plus-white responses at each wave length. The relatively high values at the spectral extremes and the minimal value in the pure yellow region are perfectly consistent both with qualitative reports and with the experimental data on this problem (e.g., 28). Again, as in the hue functions, the rate of change of the saturation coefficient from one spectral region to the next is indicative of a varying discriminative capacity with respect to wave length; and, again, the form of the function as shown applies to a moderate luminance level and varies in a determinable man-

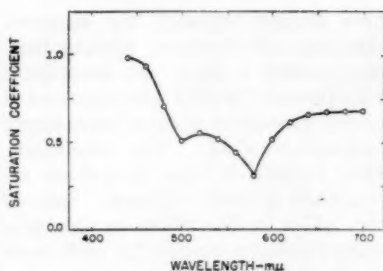


FIG. 10. Spectral saturation coefficients. Moderate luminance.

ner with change in the level of excitation in accordance with the different energy-vs.-response rates of the three independent response systems.

In view of the variations in the hue and saturation functions with change in luminance, we should expect that discrimination functions that depend on changes in these two color attributes, such as discrimination of one wave length from the next in an equal brightness spectrum, would also reflect such a dependence on luminance. Figure 11 shows, in the upper half, a set of wave-length discrimination functions obtained at two luminance levels by Weale (45).

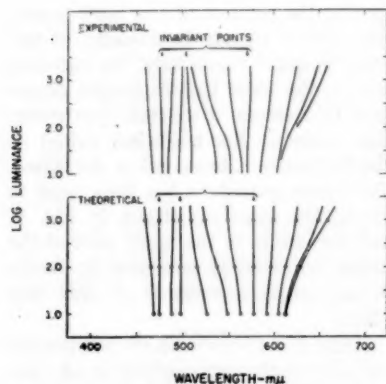


FIG. 9. Constant hue contours as measured by Purdy (38) and as predicted by theory.

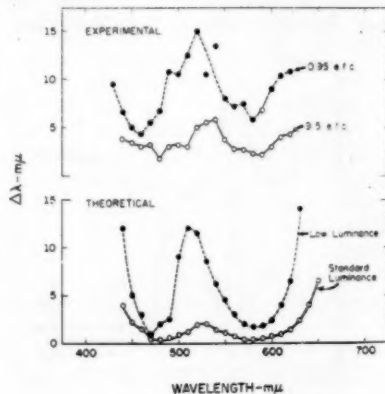


FIG. 11. Wave-length discrimination functions for two luminance levels as measured by Weale (45) and as predicted by theory.

The higher values of difference threshold obtained at the low luminance level may be explained by a general reduction of discriminative capacity in dim light. The shift of the midspectral maximum toward lower wave lengths, and the relatively greater heightening of the minimum in the yellow region, cannot, however, be attributed to such a generalized reduction in discriminatory capacity. The selectively greater loss in yellow and blue responses at the low-luminance level that is one of the postulates of our model does, however, account for changes of exactly this sort in the form of the function. This is shown by the two theoretical functions computed from pairs of spectral hue and saturation functions that are associated with the two specified luminance levels. Since brightness is kept constant in such experiments, only the hue and saturation variables need be considered in our analysis of these functions (22).

### *Chromatic Adaptation*

The phenomena that we have treated thus far all refer to the individual with normal color vision in a neutral state of adaptation. What of his color perception after the visual system has been exposed for some time to a strongly colored illuminant? For analytical purposes, the simplest situation of this sort is the one in which the eye has been exposed to a large surround field of given color and luminance, and the test stimuli are viewed at the same level of luminance as the surround. Under these circumstances, the three photochemical receptor substances will probably have undergone some selective bleaching, and because of the similar brightness of the surround and test fields, spatial induction effects in the neural response processes will probably be fairly constant. To simplify the treatment for these particular conditions, therefore, we may ignore the constant neural inductions

and consider the photosensitive changes as exercising a controlling influence on the response systems.

We know that under these circumstances the color-mixture data do not change. That is, with uniform chromatic adaptation, any change in the perceived color of one side of a bipartite color-mixture field will also occur on the other side, and to exactly the same extent. Thus a color equation that has been made with the eye adapted to a neutral white light will also be a valid equation when the eye is adapted to a colored illuminant (15). These important constancies of color equations mean that whatever photochemical changes occur with adaptation must occur in a very specific way. That is, the spectral distribution functions representing the three selective photochemicals may be selectively multiplied or reduced by a constant factor, but no one of them can change its form (44, pp. 211-212). In other words, any single substance cannot lose a greater percentage of its absorption at one wave length than it loses at another wave length. Thus, exposure to a colored light can cause any one of the postulated photochemical functions shown in Fig. 7 to be multiplied or divided by a constant amount, but this is the only alteration in the photosensitive functions that is consistent with the fact that color equations are invariant with chromatic adaptation.

The extent to which the three substances are selectively attenuated as a result of exposure to colored light is clearly controlled by the light stimulus itself. That substance which initially absorbs most of the adapting light will suffer the greatest relative bleaching, and the substance which absorbs relatively little of the adapting light will be relatively little affected by it. Thus, by determining their relative absorptions of the adapting light, we can compute the relative changes in the heights



of the photosensitive distribution functions for the three photopigments that we have postulated. Since the excitations of the opponent response systems depend on these photochemical light absorptions (see Fig. 2), we can now also determine the forms and magnitudes of the chromatic and achromatic response functions for the new condition of adaptation. In spite of the close overlap of the photosensitive functions that we have postulated, the "adapted" chromatic response functions determined in this way change in striking fashion relative to the functions for the neutral adaptation condition. The achromatic function changes too, but relatively very little. These theoretically computed adaptation changes are consistent with the kinds of change known to occur in situations of this sort. If the eye that has been adapted to white light is exposed for some time to a saturated red equal in brightness to the white, the normally red end of the spectrum does not become excessively dark, but the amount of redness seen is strongly reduced, and the greens become greatly supersaturated (3, pp. 133-137). Also, the wave length that formerly appeared pure yellow is now strongly greenish, and this is also true for the wave length that formerly appeared pure blue. These changes can be determined from the functions shown in Fig. 12 that have been computed for a given red adaptation, in comparison with the functions for the neutral state that were given in Fig. 6.

From this new set of "adapted" opponent response functions the hue and saturation coefficients and the discrimination data for this new state can also now be determined (26).

These "adapted" response functions are specified, as we said above, for a circumscribed set of conditions for which the photochemical adaptation changes could be taken as primary. As

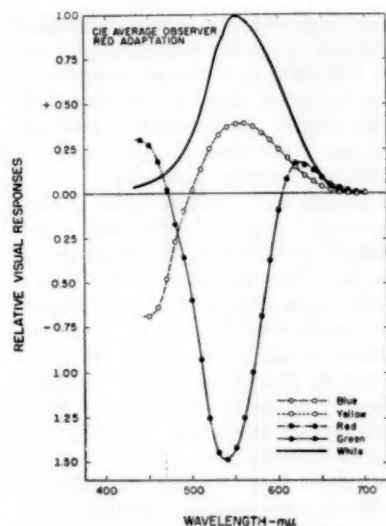


FIG. 12. Chromatic and achromatic visual response functions for red adaptation as predicted by theory.

soon as the relative luminance conditions are altered, however, then the neural inductions enter importantly into the complex picture. For example, if a test stimulus seen within a strongly colored (say, red) surround looks neutral when its luminance is the same as that of the surround, then it will probably appear somewhat reddish at a higher luminance, and the complementary green at a lower luminance (12). (The test stimulus is assumed also to be of predominantly long wave length composition.) In terms of opponent inductions this phenomenon is readily understood. If the red process excited by the red surround induces an opponent green process in the test area, then at an intermediate luminance this green induction is just strong enough to cancel the red-process activity aroused by the test stimulus itself. When the test stimulus is made brighter and the red response to it increases, the unchanged

green induction from the surround becomes inadequate to cancel completely the increased red response to the stronger test stimulus, and the red test hue is now seen. At a much lower luminance of test stimulus, the red process is activated to a much lesser extent, and the green induction from the surround, which is still unchanged in strength, is now sufficient to impart a green hue to the perceived test area. These phenomena are not only consistent with the opponent induction postulate, but they also make it clear why attempts to treat the problem of chromatic adaptation exclusively as a matter of photochemical bleaching are foredoomed to failure (e.g., 1, 33).

#### *Color Anomalies and Color Blindness*

When we come to consider individuals who do not have normal color vision we find that their color vision can depart from the normal in two general ways. Their color perceptions may be distorted relative to the normal, or they may exhibit specific color weaknesses or losses. Also, they may show both types of these deviant characteristics at the same time. By distorted color perceptions we mean, for example, the perceptions of the particular type of anomalous individual who has the following characteristics: he sees a distinct orange in the spectral region described normally as pure yellow or nearly so; he needs three stimuli for color mixture; he makes color matches with high precision but uses quite different proportions of the mixture stimuli than does the normal observer. An individual of this type does not seem to have lost any of the efficiency of his neural visual response processes, and it seems reasonable to assume that his color distortions have their basis in the photochemical complex responsible for selective light absorption.

The particular assumptions that we

have made concerning the kinds of deviation that the photosensitive materials may exhibit stem from a generalization made by Dartnall (2), on the basis of his researches concerned with the identification of visual photopigments in a variety of lower organisms. Dartnall has found that when the absorption curves of the various visual pigments are plotted as a function of the vibration frequency of the incident light (the reciprocal of the more usual wave-length specification), all the absorption curves have very nearly the same shape, and they can be made to coincide simply by shifting the curves so that they all reach an absorption maximum at the same frequency. In other words, a single template representing amount of absorption as ordinate, against frequency of radiant energy as abscissa, can be used to fit the absorption function of any visual pigment, whatever the locus of its absorption maximum. It seems reasonable to expect that this same generalization will apply to the photosensitive distributions of anomalous individuals with respect to the population of observers with normal color responses. We have consequently assumed that, in congenital abnormalities of the visual system, the normal photopigments can undergo changes that result in a uniform shift of the entire set of photosensitive distribution functions as a group along the frequency scale. These shifts are assumed to occur in either of two directions: toward higher frequencies (shorter wave lengths) resulting in the type of anomalous color vision identified as *protanomaly*, or toward lower frequencies (longer wave lengths) relative to the normal absorption loci, resulting in the second major type of anomalous color vision known as *deutanomaly*. The amount of these displacements may also vary in different degrees of congenital anomaly.

Since the absorption of light by the photosensitive materials provides the stimulus for the neural chromatic and achromatic response systems, the visual response functions thus controlled by the deviant photosensitive materials will necessarily be altered, too, and in a systematic manner. Examples of theoretically derived anomalous response functions based on these assumptions are given in Fig. 13. The set of functions in the center block are those for the observer with normal photosensitive materials; those in the upper block are for a protanomalous type whose visual pigment absorptions are assumed to be shifted toward the shorter wave lengths by an amount equal to about 15  $m\mu$

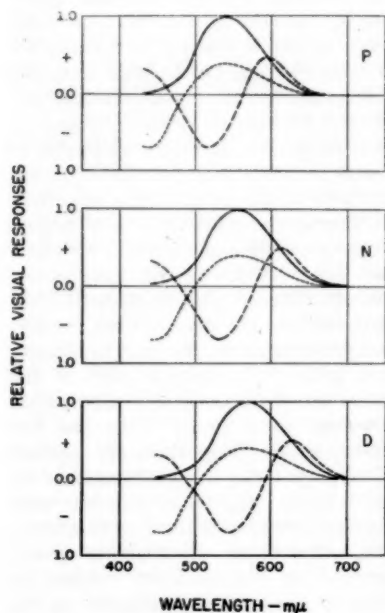


FIG. 13. Theoretical chromatic and achromatic response functions for equal energy spectrum. For observers with protanomalous, normal, and deutanomalous photoreceptor systems and with normal strength visual response processes.

from the normal peak of about 550  $m\mu$ . This type of individual will have a luminosity function (described by the achromatic, white response function) that peaks at a shorter wave length than the normal and will show considerable loss of luminosity at the red end of the spectrum (48, Ch. 25). The spectral hues will also be altered, with a distinctly reddish yellow occurring where the normal individual sees a unique or pure yellow, whereas the protanomalous observer's pure yellow occurs at a wave length described by the normal as quite greenish. In making color matches, such as a match between 589  $m\mu$  on one side of a bipartite field and a mixture of 530  $m\mu$  and 670  $m\mu$  on the other, this observer will require a much greater proportion of 670  $m\mu$  in the mixture than will the average observer with normal color vision (27, 46). This particular match, the Rayleigh equation, is the earliest and best known diagnostic test for anomalous color vision. In this same test, the anomalous individual whose response functions are shown in the lower block in Fig. 13 will deviate from the normal in the opposite way; that is, he will require a much greater proportion of 530  $m\mu$  in the mixture for the Rayleigh equation (46). This type of anomalous individual (deutanomalous) is assumed to have photopigment absorptions that are shifted toward the longer wave lengths, and he will see greenish-yellows where the normal sees yellow, yellows where the normal sees orange, etc. Since the neural response processes of both types of anomalies of this sort are assumed to be operating at the normal efficiency, these individuals will show high precision in making their distorted color matches, and their discriminatory capacities will also be good. As a matter of fact, anomalous individuals of this sort have understandably high confi-

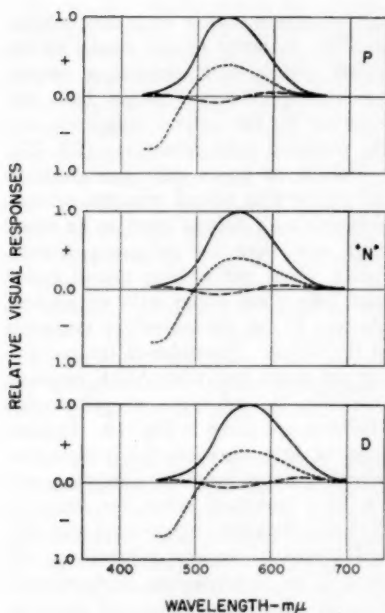


FIG. 14. Theoretical chromatic and achromatic response functions for equal energy spectrum. For observers with protanomaly, normal, and deuteranomaly photoreceptor systems, and with impaired red-green response processes.

dence in their own color capability, and they are extremely resistant toward accepting the results of diagnostic tests which indicate that their visual capacities are deviant from (with the implication of "inferior to") those of the normal population (36, pp. 235-238).

Not all anomalous individuals are as fortunate as the types shown in Fig. 13, however. Many give evidence of real color weakness, in addition to distortions of the kinds already discussed (40). These color-weak individuals seem to have been deprived of some of the efficiency of the neural response processes, particularly of the red-green opponent pair, and their systems may

be represented in terms of the theory by the kinds of response functions given as examples in Fig. 14. The visual pigments of these three types of individuals are taken to be the same as those shown in the preceding figure, respectively, but the red-green paired system is reduced to one-tenth of the normal strength. Such observers have real losses in color discrimination in addition to possible color distortions, and their color matches are imprecise as well as deviant. Individuals with congenitally abnormal color systems are frequently of this general type, and cases of acquired color blindness caused by degenerative disease invariably show this kind of color weakness at some stage in the development of the neural disorder (31).

When the weaknesses become extreme, whether in congenital or acquired disorders, the red-green system seems to be entirely lost to normal function, and a condition of dichromasy, or so-called "color-blindness," results. That is, the visual system becomes a two-variable

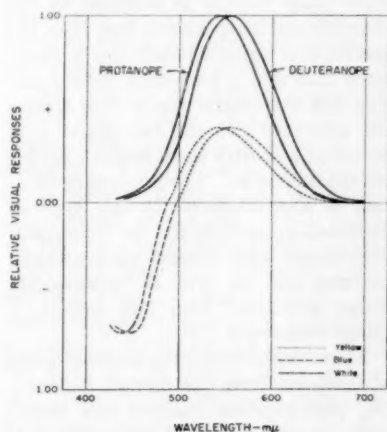


FIG. 15. Theoretical chromatic and achromatic response functions for equal energy spectrum. For observers with nonfunctioning red-green response processes.

one, as shown in Fig. 15. Here the yellow-blue and the white-black neural systems remain intact and functioning, but there is no red-green response function. If the red-green loss occurs without changes in the visual pigments, the remaining yellow-blue and white-black response functions are like those of the normal individual; but, since there is no red-green system, the spectrum is divided into only two hue sections for these individuals. The short wave lengths which normally vary from violet through blue and blue-green to pure green all appear as blue, but of varying saturations, with a neutral region where the normal pure green occurs. Beyond this wave length the remainder of the spectrum appears yellow, in varying saturations, out to the extreme long-wave limit of visibility. The luminosity function is the same as for the observer with normal color vision. Individuals who fit this response pattern would be classified as *deuteranopes* (29). If the visual pigments are altered, so as to produce an absorption shift toward the short wave lengths in addition to the complete red-green neural loss, then the spectrum is again divided into a short-wave blue and a long-wave yellow section, but the neutral region that divides the spectrum into the two major hues occurs at a shorter wave length than for the deuteranopes. The luminosity function is also displaced in this type of dichromasy, as it is for the anomalous individuals with similar photopigment changes, and the type of "color-blind" vision associated with this pattern is called *protanopia* (29).

These two theoretically assumed kinds of deviation from the normal system—i.e., photopigment changes and neural losses or weaknesses of the paired red-green response system—permit us to assemble a systematic picture of the many various manifestations of abnor-

mal red-green vision that defy understanding in terms of any model of the visual system that assumes a one-to-one correspondence between light absorption in the retinal receptors and the resulting color sensations (22, 27).

Defects or losses may also occur in the yellow-blue neural response system, although such defects seem to be much more rare than the red-green defects. Again, these yellow-blue neural losses may take place either with or without changes in the photosensitive materials in the retina. Examples of the remaining red-green and white-black response functions in two types of yellow-blue blindness are given in Fig. 16. In each type of this disorder, the yellow-blue neural response function is missing, and the total gamut of colors for these individuals includes only neutral and reds and greens of various saturations. If there is no simultaneous photopigment disorder, there are two neutral points in the spectrum, one in the region where the normal sees a pure yellow, and an-

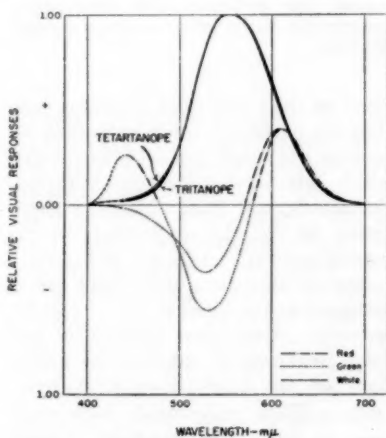


FIG. 16. Theoretical chromatic and achromatic response functions for equal energy spectrum. For observers with nonfunctioning yellow-blue response processes.



other in the region where the normal sees a pure blue. Yellow-blue blindness of this sort is called *tetartanopia*, and only a few cases of it have been reported in the literature (e.g., 34, pp. 68-92). Slightly more common is the second type of yellow-blue blindness, known as *tritanopia* (49), in which not only the neural yellow-blue system is lost, but also the short-wave photopigment seems to be missing. Observers of this type have a neutral point in the normally yellow-green region of the spectrum, but there is no second neutral point, and the green hues extend into the short-wave region that appears violet to the person with normal color vision.

For all these types of deviant color vision, calculation from the theoretical spectral response functions of discrimination curves, color mixture equations, and other psychophysical relations are in good agreement with the experimental data that are available for the various kinds of defective color systems (22, 27).

#### *Opponents-Theory and Neurophysiology*

The conceptual model for the opponent-colors theory as originally presented by Hering drew its sharpest criticism on the grounds of being bad physiology. Some of this criticism was based on an erroneous interpretation of Hering's views, an interpretation that incorrectly assigned the opponent processes to the photochemical activities in the retinal cells. Hering's own concept of mutually opponent neural processes, each capable of being activated by external stimulation, was also, however, far ahead of the knowledge of neurophysiology at the time it was proposed (16). But this concept now turns out to be perfectly consistent with the picture of neural function that is only just recently beginning to build up from elec-

trophysiological studies of the visual neural apparatus.

It has become clear that nerves do not simply respond or fail to respond when a stimulus is presented to the appropriate end-organ. Rather, they may respond according to any of a number of quite specific patterns. For example, a nerve fiber may (a) discharge at the onset of stimulation and subsequently gradually become quiet; (b) discharge at both onset and cessation of stimulation with a quiet period in between; or (c) cease any spontaneous activity when first stimulated and during continued stimulation, but respond with a burst of electrical impulses when the stimulus ceases to act (7). The on- and off-phases of discharge are mutually inhibitory processes, they are associated with slow electrical potentials of opposite sign, and they cancel each other when the experimental conditions are so manipulated as to cause both on- and off-discharges to impinge simultaneously on the same ganglion cell (6). In Granit's opinion (6), the evidence from electrophysiology provides a "belated vindication of Hering's view" that the visual system is characterized by mutually opponent neural processes.

The concept of mutual interaction among the various elements of the physiological field is also basic to the theory and is critical to an understanding of both areal effects and simultaneous contrast phenomena. Here again, we find the researches in electrophysiology indicating that individual nerve elements never act independently, and that visual function must be thought of in terms of the integrated action of all the units of the neural visual system (8). Hartline (9) has found that, even in the very simple *Limulus* eye, the discharge of impulses in any one optic nerve fiber depends not only upon the stimulus to the specific receptor unit

from which that fiber arises but also upon the stimulation over the entire population of mutually interacting elements. Both excitatory and inhibitory interactions of the sort to be expected by theory have actually been demonstrated in the neural responses of the vertebrate visual system by Hartline (8), Kuffler (32), and Granit (6).

The way in which the postulated three independent systems of paired opponent processes (*y-b*, *r-g*, *w-bk*) are differentiated neurally is still a matter for conjecture. Hering thought it was a matter of process specificity, but was willing to use the concept of material, or structural, specificity, which he guessed would be more readily comprehended by most interested readers of his views at the time. Our own theoretical preference at this time is the conjecture that a particular color quality is more probably determined by a particular state of the nervous tissue than by activity of a particular structural element in the nervous network. Thus, we would be inclined to look for a difference between yellow-blue vs. red-green processes, rather than toward isolation of yellow-blue or red-green fibers or nerve cells.

#### SUMMARY

This paper has presented a summary of our progress to date in providing a quantitative formulation for the Hering opponent-colors theory, and in relating the postulated visual mechanism to specific problems of color sensation, color mixture and color discrimination; to the dependence of these functions on the physical variables of both stimulus wave length and energy level; to their further dependence on adapting and surround stimulation; and to the changes in these functions that occur in various kinds of abnormal color vision. It is

our conclusion that the opponent-colors theory serves as a fruitful working hypothesis by bringing a systematic coherence to the mass of isolated color phenomena that have been reported and subjected to quantitative experiment throughout the years. The physiological concepts basic to the theory are also shown to be consistent with recent findings in neurophysiology.

#### REFERENCES

1. BREWER, W. L. Fundamental response functions and binocular color matching. *J. Opt. Soc. Amer.*, 1954, **44**, 207-212.
2. DARTNALL, H. J. A. The interpretation of spectral sensitivity curves. *Brit. med. Bull.*, 1953, **9**, 24-30.
3. EVANS, R. M. *An introduction to color*. New York: Wiley, 1948.
4. FARNSWORTH, D. Tritanomalous vision as a threshold function. *Die Farbe*, 1955, **4**, 185-196.
5. GRAHAM, C. H. Vision: III. Some neural correlations. In C. Murchison (Ed.), *A handbook of general experimental psychology*. Worcester: Clark Univer. Press, 1934. Pp. 829-879.
6. GRANIT, R. *Receptors and sensory perception*. New Haven: Yale Univer. Press, 1955.
7. HARTLINE, H. K. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *Amer. J. Physiol.*, 1938, **121**, 400-415.
8. HARTLINE, H. K. The neural mechanisms of vision. *Harvey Lectures*, 1941-42, **37**, 39-68.
9. HARTLINE, H. K., WAGNER, H. G., & RATLIFF, F. Inhibition in the eye of limulus. *J. gen. Physiol.*, 1956, **39**, 651-673.
10. HARTRIDGE, H. The polychromatic theory. *Documenta Ophthalm.*, 1949, **3**, 166-193.
11. HECHT, S. Vision: II. The nature of the photoreceptor process. In C. Murchison (Ed.), *A handbook of general experimental psychology*. Worcester: Clark Univer. Press, 1934. Pp. 704-828.
12. HELSON, H. Fundamental problems in color vision. I. The principle governing changes in hue, saturation, and lightness of non-selective samples in

- chromatic illumination. *J. exp. Psychol.*, 1938, 23, 439-476.
13. HERING, E. *Zur Lehre vom Lichtsinne*. Berlin, 1878.
14. HERING, E. Zur Erklärung der Farbenblindheit aus der Theorie der Gegenfarben. *Lotos, Jb. f. Naturwiss.*, 1880, 1, 76-107.
15. HERING, E. Ueber Newton's Gesetz der Farbmischung. *Lotos, Jb. f. Naturwiss.*, 1887, 7, 177-268.
16. HERING, E. *Zur Theorie der Vorgänge in der lebendigen Substanz*. Prague: 1888. (English translation by F. Welby, in *Brain*, 1897, 20, 232-258.)
17. HERING, E. *Grundzüge der Lehre vom Lichtsinn*. Berlin: Springer, 1920.
18. HOCHBERG, J. E., TRIEBEL, W., & SEAMAN, G. Color adaptation under conditions of homogeneous visual stimulation (Ganzfeld). *J. exp. Psychol.*, 1951, 41, 153-159.
19. HURVICH, L. M., & JAMESON, DOROTHEA. The binocular fusion of yellow in relation to color theories. *Science*, 1951, 114, 199-202.
20. HURVICH, L. M., & JAMESON, DOROTHEA. Spectral sensitivity of the fovea. I. Neutral adaptation. *J. Opt. Soc. Amer.*, 1953, 43, 485-494.
21. HURVICH, L. M., & JAMESON, DOROTHEA. A quantitative theoretical account of color vision. *Trans. N. Y. Acad. Sci.*, 1955, 18, 33-38.
22. HURVICH, L. M., & JAMESON, DOROTHEA. Some quantitative aspects of an opponent-colors theory. II. Brightness, saturation, and hue in normal and dichromatic vision. *J. Opt. Soc. Amer.*, 1955, 45, 602-616.
23. HURVICH, L. M., & JAMESON, DOROTHEA. Some quantitative aspects of an opponent-colors theory. IV. A psychological color specification system. *J. Opt. Soc. Amer.*, 1956, 46, 416-421.
24. ISHAK, I. G. H. Determination of the tristimulus values of the spectrum for eight Egyptian observers and one British observer. *J. Opt. Soc. Amer.*, 1952, 42, 844-849.
25. JAMESON, DOROTHEA, & HURVICH, L. M. Some quantitative aspects of an opponent-colors theory. I. Chromatic responses and spectral saturation. *J. Opt. Soc. Amer.*, 1955, 45, 546-552.
26. JAMESON, DOROTHEA, & HURVICH, L. M. Some quantitative aspects of an opponent-colors theory. III. Changes in brightness, saturation, and hue with chromatic adaptation. *J. Opt. Soc. Amer.*, 1956, 46, 405-415.
27. JAMESON, DOROTHEA, & HURVICH, L. M. Theoretical analysis of anomalous color vision. *J. Opt. Soc. Amer.*, 1956, 46, 1075-1089.
28. JONES, L. A., & LOWRY, E. M. Retinal sensibility to saturation differences. *J. Opt. Soc. Amer.*, 1926, 13, 25-34.
29. JUDD, D. B. Current views on colour blindness. *Documenta Ophthalmol.*, 1949, 3, 251-288.
30. JUDD, D. B. Basic correlates of the visual stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 811-867.
31. KÖLLNER, H. *Die Störungen des Farbensinnes*. Berlin: S. Karger, 1912.
32. KUFFLER, S. W. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.*, 1953, 16, 37-68.
33. MACADAM, D. L. Chromatic adaptation. *J. Opt. Soc. Amer.*, 1956, 46, 500-513.
34. MÜLLER, G. E. *Darstellung und Erklärung der verschiedenen Typen der Farbenblindheit*. Göttingen: Vandenhoeck and Ruprecht, 1924.
35. PARSONS, J. H. *An introduction to the study of colour vision*. (2nd ed.) Cambridge: Cambridge Univ. Press, 1924.
36. PICKFORD, R. W. *Individual differences in colour vision*. London: Routledge and Kegan Paul, 1951.
37. PIÉRON, H. La dissociation de l'adaptation lumineuse et de l'adaptation chromatique et ses conséquences théoriques. *Année psychol.*, 1939, 40, 1-14.
38. PURDY, D. M. The Bezold-Brücke phenomenon and contours for constant hue. *Amer. J. Psychol.*, 1937, 49, 313-315.
39. RIGGS, L. A., RATLIFF, F., CORNSWEET, JANET C., & CORNSWEET, T. N. The disappearance of steadily fixated visual test objects. *J. Opt. Soc. Amer.*, 1953, 43, 495-501.
40. ROSMANIT, J. *Anleitung zur Feststellung der Farbentüchtigkeit*. Leipzig: Deuticke, 1914.
41. TROLAND, L. T. Apparent brightness: its conditions and properties. *Trans. Illum. Engr. Soc.*, 1916, 11, 957-966.
42. TROLAND, L. T. *The principles of psychophysiology*. Vol. 2. *Sensation*. New York: D. Van Nostrand, 1930.

43. VON BEZOLD, W. Ueber das Gesetz der Farbenmischung und die physiologischen Grundfarben. *Ann. Phys. u. Chem.*, 1873, **150**, 221-247.
44. VON KRIES, J. Die Gesichtsempfindungen. In W. Nagel (Ed.), *Handbuch der Physiologie des Menschen*. Brunswick: Vieweg, 1905. Pp. 109-282.
45. WEALE, R. A. Hue-discrimination in para-central parts of the human retina measured at different luminance levels. *J. Physiol.*, 1951, **113**, 115-122.
46. WILLIS, MARION P., & FARNSWORTH, D. Comparative evaluation of anomalous scopes. *Med. Res. Lab. Rep.* No. 190, 1952, **11**, No. 7, 1-89.
47. WRIGHT, W. D., & PITT, F. H. G. The colour-vision characteristics of two trichromats. *Proc. Phys. Soc. (London)*, 1935, **47**, 205-217.
48. WRIGHT, W. D. *Researches on normal and defective colour vision*. St. Louis: Mosby, 1947.
49. WRIGHT, W. D. The characteristics of tritanopia. *J. Opt. Soc. Amer.*, 1952, **42**, 509-521.

(Received for early publication July 1, 1957)

**ANNOUNCEMENT**

**JOURNAL OF EDUCATIONAL  
PSYCHOLOGY**

This journal will now be published by the American Psychological Association. In 1958 it will become a bimonthly; issues will appear in February, April, June, August, October, and December. Contents include articles on problems of teaching, learning, and the measurement of psychological development.

All back issues and subscriptions up to and including the May 1957 issue are the property of Warwick and York, Inc., 10 East Centre Street, Baltimore 2, Maryland.

Subscription \$8.00  
(Foreign, \$8.50)

Single  
Copies, \$1.50

*Direct new subscriptions and renewals to:*

**AMERICAN PSYCHOLOGICAL ASSOCIATION**  
Publications Office  
1333 Sixteenth Street, N. W.  
Washington 6, D. C.



# Important Evergreen Psychology Books \$1 and up

The same fine soft-cover editions used in college psychology courses throughout the country. Now at your bookstore.

- Masochism in Modern Man** by Theodor Reik. \$1.95
- A Study of Interpersonal Relations.** New Contributions to Psychiatry, edited by Patrick Mullaney. \$1.95
- Psychoanalysis: Evolution and Development.** A Review of Theory and Therapy, by Clara Thompson and Patrick Mullaney. \$1.45
- Black Anger.** A unique psychoanalytic biography of a native African medicine man, by Wail Sacks. \$1.75
- Captain Dreyfus.** The Story of a Man Hysteria, by Nicholas Holms. \$1.45
- The Forgotten Language.** An Introduction to the Understanding of Dreams, Fairy-tales and Myths, by Erich Fromm. \$1.75
- Rebel Without a Cause.** The hypnosis of a criminal psychopath, by Robert M. Lindner. \$1.45
- The Bisexual Nature of Man** by Ashley Montagu. \$1.00
- Listening with the Third Ear.** The Inner Experience of a Psychiatrist by Theodor Reik. \$1.95
- Democracy and Dictatorship.** An incisive study of the psychological and social conditions that produce each, by Zaveloff Berke. \$1.45
- Jung's Psychology and Its Social Meaning** by Ira Prosser. \$1.25
- Oedipus: Myth and Complex.** A review of all major psychoanalytic theories by Patrick Mullaney. \$1.45
- Immortality.** Its psychological, biological and anthropological implications, by Ashley Montagu. \$1.00
- The Marquis de Sade.** An essay by Simone de Beauvoir, with selections from De Sade's writings. \$1.45

SEND FOR FREE DESCRIPTIVE CATALOG

**grove press**  
795 Broadway, New York 3